

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Zemědělská fakulta

**Využití statistických metod při analýze  
dat plynoucích z biologických experimentů**

**Habilitační práce**

2019

Ing. Michael Rost, Ph.D.

Prohlašuji, že jsem habilitační práci na téma „**Využití statistických metod při analýze dat plynoucích z biologických experimentů**“ zpracoval samostatně a s využitím vlastních publikovaných výsledků, které jsou nedílnou součástí této práce. K sepsání předkládané práce byly dále použity citované publikace uvedené v seznamu použité literatury.

V Českých Budějovicích dne 5. 9. 2019

Ing. Michael Rost, Ph.D.

Rád bych na tomto místě poděkoval především svým rodičům a své dceři za dlouhodobou podporu. Rovněž bych chtěl poděkovat kolegům z Katedry aplikované matematiky a informatiky při EF JU, jíž jsem byl více jak dvanáct let členem. Dále celé radě spolupracovníků ze ZF a FROV při JU v Českých Budějovicích, bez kterých by nikdy nevznikla řada publikací, na kterých jsem se podílel.

Michael Rost

# Obsah

<b>1 Úvod</b>	<b>5</b>
1.1 Členění práce . . . . .	6
1.2 Cíl práce . . . . .	7
1.3 Publikační výstupy s přímou vazbou na habilitační práci . . .	8
<b>2 Hodnocení experimentálních dat</b>	<b>10</b>
2.1 Návrh a analýza experimentů . . . . .	10
2.2 Klasické přístupy k hodnocení experimentálních dat . . . . .	15
2.2.1 Základní charakteristiky a příbuzné deskriptivní nástroje . . . . .	15
2.2.2 Tradiční konfirmační metody . . . . .	16
2.2.3 Klasický přístup k modelování závislostí . . . . .	17
2.2.4 Klasické metody vícerozměrné analýzy dat . . . . .	19
2.3 Některé moderní přístupy k hodnocení experimentálních dat	23
2.4 Další směřování ve využití statistických metod v experimentální praxi . . . . .	28
<b>3 Komentář k publikovaným pracím</b>	<b>30</b>
<b>4 Závěr a přínos habilitační práce pro obor</b>	<b>40</b>

5 Implementace dosažených poznatků ve vysokoškolské výuce	42
6 Publikační činnost autora	44
7 Publikované práce s přímou vazbou na habilitační práci	48
Literatura . . . . .	156

# Úvod

Je zcela zřejmé, že se při vyhodnocování výsledků plynoucích z observačních studií, či provedených experimentů, neobejdeme bez objektivního přístupu k jejich následnému vyhodnocení a interpretaci. Právě zde hraje svou úlohu aplikovaná statistika a její rozmanitý aparát, který umožňuje výsledky zobecňovat a současně s využitím pravděpodobnostního přístupu i kvantifikovat, resp. odhadnout a minimalizovat chyby, či nepřesnosti, jichž se můžeme při statistické indukci dopustit.

S rostoucími požadavky na kvalitu experimentální práce rostou nároky na zpracování získaných dat. V dnešní době je takřka nepředstavitelné, aby jeden člověk dokázal navrhnout, provést a zpracovat komplexní experiment zahrnující různé metodiky či moderní diagnostické techniky, které jsou založeny na různých vědních disciplínách. Ideálním případem je, aby kromě biologů pracoval na přípravě a vyhodnocení experimentů specialista, statistik.

Existuje celá řada statistických metod, které umožňují hodnocení a zpracování výsledků, počínaje aplikací klasických metod, přes vícerozměrné metody až po funkcionální analýzu dat.

Správné naplánování experimentu a použití vhodných statistických metod umožňuje získání mnohdy ne zcela zřejmých výsledků, kterých by nebylo dosaženo bez spolupráce experimentátora se statistikem. Snahou předložené práce je poskytnout právě takový přehled vybraných statistických metod aplikovatelných v různých fázích experimentální a vědecké práce spolu s jejich využitelností pro hodnocení experimentálních výsledků.

## 1.1 Členění práce

Předložená habilitační práce je členěna do několika partií, dle jejich převažujícího zaměření.

První část práce pojednává o metodologii plánování experimentů<sup>1</sup>, kterou by mělo vše začínat a která by měla být východiskem vědecké práce. Pozornost je věnována zejména problematice faktoriálních plánů a tzv. simplexních plánů pro návrhy směsí spolu s přesahem do následné metodologie analýzy responsních povrchů<sup>2</sup>, kterou autor uplatnil při několikaleté práci, která vyústila v podání přihlášky evropského patentu číslo EP18196664.9 dne 25. 9. 2018 s názvem „Pharmaceutical Composition Comprising Solid Dispersions of Amorphous Cladribine and Pharmaceutically Acceptable Water Soluble Carrier“. Přihláška bude dle platných zákonných procedur vyložena na konci roku 2019. Následně jsou v předložené práci zmíněny i optimalizační přístupy k dosažení co možná nejlepší response - odezvy v podobě minima, maxima, či předem definované cílové hodnoty. Některé z těchto technik jsou ve své podstatě heuristiky, které jsou inspirovány přírodou.

Druhá část poskytuje přehled některých klasických statistických metod a jejich využití při vyhodnocování experimentálních dat. Zmíněny jsou zde i další klasické, avšak ne příliš využívané možnosti základního zpracování dat. Následně jsou zmíněny tradiční přístupy, využívané pro komparativní účely plánovaných experimentů.

V této partii jsou dále zmíněny možnosti modelování vztahů mezi proměnnými. Jde především o přehled některých přístupů k řešení různých „regresních úloh“, konkrétněji použití lineárních regresních modelů, lineárních smíšených modelů, nelineární regresních modelů a zobecněných lineárních modelů, zejména logistické regrese. Součástí druhé partie je i přehled některých tradičních vícerozměrných metod. Zmiňovány jsou metody shlukové analýzy, analýza hlavních komponent, metody diskriminační analýzy. Krátce je diskutována metodologie klasifikačních a regresních stromů.

Třetí část práce přináší přehled vybraných statistických metod, které lze považovat za modernější analogy metod popsanych ve druhé části práce.

---

<sup>1</sup>DOE - z angl. Design of Experiments

<sup>2</sup>RSM - z angl. Response Surface Methodology

Představeny jsou novější klasifikační metody, které lze považovat za rozšíření výše diskutovaných metod jako je např. regularizovaná diskriminační analýza nebo metoda náhodných lesů založená na výhodě získané souborným hlasováním. Diskutována je i problematika funkcionální analýzy dat. Rovněž jsou zmíněny i metody využitelné při moderní high-throughput analýze u tzv. „-omic“ dat, které se v současnosti prosazují v experimentální praxi.

Dostupnost výše zmiňovaných metod je mnohdy svázaná s příslušným softwarovým vybavením. V této souvislosti je v práci na řadě míst zmiňováno především freewarové softwarové prostředí R [111], které je dnes ve své podstatě možno chápat za jakousi „Lingua franca“ odborné veřejnosti věnující se aplikované statistice.

## 1.2 Cíl práce

Předložená habilitační práce si klade za cíl:

- a) Podat přehled statistických metod využitelných při analýze dat plynoucích z experimentální činnosti či pozorování projevů živých organismů.
- b) Seznámit čtenáře s možnostmi uplatnění různých kvantitativních přístupů při plánování, analýze a interpretaci experimentálních dat. Zejména pak s metodologií plánování experimentů, modelování závislostí, zpracování vícerozměrných dat, klasifikačních metod a aparátu funkcionální analýzy dat.
- c) Stručně nastínit možný vývoj aplikace statistických metod v experimentální praxi.
- d) Na příkladech vlastních výsledků pak ukázat využitelnost níže diskutovaných statistických metod při analýze z různých experimentálních oblastí, zahrnujících různé typy experimentů (*in vivo*, *in vitro*, formulace finálních lékových forem).



### 1.3 Publikační výstupy s přímou vazbou na habilitační práci

Níže je uveden přehled publikací s přímou vazbou na předloženou habilitační práci. Odkazy, resp. citace, odpovídají zařazení v použité literatuře, která je uvedena na konci práce (odkazy jsou vyvedeny tučně).

- [54] Holubová, N., Sak, B., Hlásková, L., Květoňová, D., Hanzal, V., Rajský, D., Rost, M., McEvoy, J. Kváč, M. Host specificity and age-dependent resistance to *Cryptosporidium avium* infection in chickens, ducks and pheasants. *Experimental Parasitology*, 2018, 191, 8, 62-65. IF 1,821
- [68] Kváč, M., Kodádková, A., Sak, B., Květoňová, D., Jalovecká, M., Rost, M., Salát, J. Activated CD8+ T cells contribute to clearance of gastric *Cryptosporidium muris* infections. *Parasite Immunology*, 2011, 33, 4, 210-216. IF 2,836
- [72] Kváč, M., Němejc, K., Kestránová, M., Květoňová, D., Wágnerová, P., Kotková, M., Rost, M., Samková, E., McEvoy, J., Sak, B. Age related susceptibility of pigs to *Cryptosporidium scrofarum* infection. *Veterinary Parasitology*, 2014, 202, 3-4, 330-334. IF 2,422
- [87] Másílko, J., Hartvich, P., Rost, M., Urbánek, M., Hlaváč, D., Dvořák, P. Potential for Improvement of Common Carp Production Efficiency by Mechanical Processing of Cereal Diet. *Turkish Journal of Fisheries and Aquatic Sciences*, 2014, 14, 1, 145-153. IF 0,482
- [99] Mynářová, A., Foitová, I., Kváč, M., Květonová, D., Rost, M., Morrogh-Bernard, H., Nurcahyo, W., Nguyen, C., Supriyadi, S., Sak, B. Prevalence of *Cryptosporidium* spp., *Enterocytozoon bienersi*, *Encephalitozoon* spp. and *Giardia intestinalis* in Wild, Semi-Wild and Captive Orangutans (*Pongo abelii* and *Pongo pygmaeus*) on Sumatra and Borneo, Indonesia. *Plos ONE*, 2016, roč. 11, č.3. IF 2,766
- [102] Němejc, K, Sak, B., Květoňová, D., Kernerová, N., Rost, M., Cama, VA., Kváč, M. Occurrence of *Cryptosporidium suis* and *Cryptosporidium scrofarum* on commercial swine farms in the Czech Republic and its associations with age and husbandry practices *Parasitology Research*, 2013, 112, 3, 1143-1154. IF 2,558

- [107] Perlík, F., Masri, M.A., Rost, M., Kamarád, V. Pharmacokinetic conversion study of a new cyclosporine formulation in stable adult renal transplant recipients. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2005, 149, 2, 309-13. IF 1,141
- [119] Rutkayová, J., Vácha, F., Maršálek, M., Beneš, K., Civišová, H., Horká, P., Petrášková, E., Rost, M., Šulista, M. Fish stock losses due to extreme floods - findings from pond-based aquaculture in the Czech Republic. Journal of Flood Risk Management, 2018, 11, 3, 351-359. IF 2,483
- [135] Urbánek, M., Hartvich, P., Vácha, F., Rost, M. Investigation of fat content in market common carp (*Cyprinus carpio*) flesh during the growing season. Aquaculture nutrition, 2010, 16, 5, 511-519. IF 0.83
- [136] Vácha, F., Cepák, M., Urbánek, M., Vejsada, P., Hartvich, P., Rost, M. Impact Of Long-Term Storage On The Instrumental Textural Properties Of Frozen Common Carp (*Cyprinus Carpio*, L.) Flesh. International Journal of Food Properties, 2013, 16, 2, 241-250. IF 0,915

# Hodnocení experimentálních dat

Na tomto místě je třeba zároveň říci, že jde pouze o přehled vybraných metod se kterými se autor setkal či je případně využil ke své práci, ať již publikačního charakteru nebo při práci pro různé instituce či firmy během své profesní kariery. Zdaleka se nejedná o ucelený přehled dostupných metod soudobé aplikované statistiky, neboť ten je mnohem rozmanitější.

## 2.1 Návrh a analýza experimentů

Lze říci, že za jednu z poněkud opomíjených oblastí vědecké práce můžeme označit plánování experimentální činnosti, u které cesta za poznáním začíná. Určitě lze souhlasit s tvrzením, že pro inovace, zlepšení, získání nové znalosti je imperativem právě experimentální činnost a způsob hodnocení získaných dat. Zkrátka hledání jiného, nového přístupu, který sebou potenciálně přináší nové informace, vedoucí ke snížení entropie studovaného fenoménu. Důležitost této činnosti rovněž podtrhuje skutečnost, že v prvních fázích experimentování je zpravidla vynaloženo velké úsilí, nezdědka spojené se spotřebou větší části dostupných finančních prostředků. Ale zpět k problematice.

Pokud bychom chtěli vyslovit definici slova experiment, mohli bychom říci, že jde o určitou plánovanou, opakovatelnou a ověřitelnou lidskou činnost, jejíž cílem je prokázat, nebo vyvrátit určité tvrzení - poznatek o kauzálních vztazích studovaného fenoménu. Výsledkem experimentu je tedy, či by měla být, informace. Jak již bylo zdůrazněno v definici, důležitou vlastností experimentu je možnost jeho opakování, neboť podmínky experimentu (stav prostředí) jsou ve své podstatě výběrem z množiny možných stavů a může

dojít k tomu, že výsledek je ve finále funkcí koincidence dalších faktorů. Teprve opakováním experimentu, ať již ve stejných či pozměněných podmínkách, může ukazovat na kauzální závislosti studovaných jevů.

Při experimentální práci je třeba rovněž klást důraz na přesnost měření, pozorování a popis celého experimentu i jeho jednotlivých fází. V mnohých případech je důležitá i možnost srovnání experimentálního a kontrolního vzorku, viz partie níže.

Samotné plánování experimentální činnosti mělo však svůj historický vývoj. Asi nejpoužívanějším způsobem experimentování byl přístup, při kterém docházelo k postupné změně jednoho faktoru, zatímco ostatní faktory zůstávaly na konstantní úrovni. V anglické literatuře se pro tento postup experimentování vžilo označení OFAT [15], neboli one-factor-at-a-time.

Vzhledem k tomu, že docházelo ke změně vždy jen jednoho faktoru, byl takový způsob experimentování dosti náročný na počet experimentů a počítá se i na zdroje, čas i finanční prostředky. Mnohdy byl tento způsob experimentování dosti závislý na intuici a zkušenosti osoby, která experimenty prováděla. Nehledě na to, že často postup vedl pouze k dosažení zdánlivě optimálních podmínek studovaného procesu.

S růstem složitosti průmyslové výroby vzrůstaly i požadavky na experimentování. Přístup OFAT se stal vzhledem k náročnosti experimentů a nutnosti sledování vzájemných vztahů mezi sledovanými faktory neefektivní. Bylo tedy nutno změnit přístup k provádění experimentů. Jako vhodný se ukázal přístup, který umožňuje posuzovat všechny uvažované faktory simultánně. Dnes tento nový přístup známe pod názvem DOE, která je zkratkou pro „Design of Experiment“, tedy pro přístup k plánování experimentů popsany velmi výstižně Montgomerym v monografii [90].

Základy pro metodologii DOE rozpracoval sir Ronald Fisher ve 30. letech dvacátého století v úzké souvislosti s analýzou rozptylu při polních pokusech na Rothamstedské experimentální stanici v Harpendenu. Poznatky shrnul Fisher ve své knize „The Design of Experiment“ [39], která vyšla v roce 1935. Nelze nezmínit i jeho starší knihu „Statistical Methods for Research Workers“ [38], která se stala v poválečných letech jakýmsi standardem.

Ve své podstatě lze z určitého pohledu rozlišovat několik přístupů pro návrh experimentů šetřících čas a prostředky pomocí vhodně zvolených kombinací pozorování, resp. experimentálního nastavení, které zároveň minimalizovalo počet nutných měření.

Za dnes již klasický přístup lze označit faktoriálové experimenty [98] případně jejich části, přesněji částečně faktoriálové experimenty s různým podílem frakce a tzv. rozlišením, viz [145]. Blíže je problematika popsána např. v publikaci [13] nebo [88], ve kterých se autoři věnují např. plánům  $2^k$ ,  $2^{k-p}$  nebo  $2_{IV}^{k-p}$  pro dvouúrovňové návrhy a dalším speciálním plánům<sup>1</sup>.

Návrhy založené na faktoriálových plánech umožňují především screeningovou činnost<sup>2</sup>. Jinými slovy, umožňují prvotní identifikaci důležitých faktorů ovlivňujících responsní proměnnou. Pochopitelně existují i jiné přístupy například Burmanův-Plackettův návrh [108], supersaturované plány [81] nebo dokonce i zcela jiná metodologie, viz Taguchi [117], založená na tzv. „ztrátové funkci“ umožňující identifikaci vlivných faktorů. Jako příklady využití těchto metod lze uvést práce zabývající se např. optimalizací PCR<sup>3</sup> reakce [22] nebo práce [35] využívající Burmanův-Plackettův návrh.

Po identifikaci vlivných proměnných/faktorů většinou následuje iterační proces jehož cílem je dosažení oblasti, ve které se nalézá optimum, nebo je této oblasti alespoň co nejbližší. K tomuto účelu bývá zpravidla využívána metoda největšího spádu, využívající znalosti gradientu<sup>4</sup>, viz např. [90] nebo [138]. Následně jsou využívány návrhové plány při více úrovních [66], mající za cíl zachytit možný tvar responsního povrchu v blízkosti oblasti optima. Za tímto účelem je uplatňována RSM, viz literatura [14], [16], [67], [29] nebo na biotechnologii zaměřená monografie Haalanda [49]. Příklady použití tohoto přístupu lze nalézt v pracích [132], [147].

Nezřídka se v praxi stává, že je experimentální prostor ohraničen omezujícími podmínkami. V takových případech je nutné využít speciálních návrhových plánů [27]. Zpravidla jde o tzv. „abecedně“ optimální plány, zmiňované např. v literatuře [98] či v monografii popisující různé případové studie, viz [46]. K jejich vygenerování se využívá různých algoritmů, příkladem může být statistickým softwarem využívaný algoritmus McLeana a Andersona popsaný v článku [93].

---

<sup>1</sup>Například tzv. supersaturovaným plánům. Pozn.: V případě supersaturovaných plánů je počet experimentů menší než počet sledovaných faktorů a v některých případech neumožňují návrhy odhadnout všechny hlavní efekty. Související teorie je popsána např. v publikacích [81], [82] nebo [45]

<sup>2</sup>Mezi screeningové plány patří i určitá množina směsných plánů, viz níže. V případě těchto návrhů nejsou jednotlivé proměnné nezávislé.

<sup>3</sup>Polymerase Chain Reaction

<sup>4</sup>Níže v textu budou zmíněny i jiné, modernější přístupy založené na nederivačních metodách

V praxi se však můžeme setkat i se situací kdy je výsledná hodnota response ovlivněná nikoliv úrovněmi jednotlivých faktorů, ale spíše jejich vzájemným poměrem [24], [21]. V takovém případě nejsou jednotlivé faktory/složky/proměnné nezávislé. Jinými slovy, návrhy musejí splňovat omezující podmínku  $X_1 + X_2 + \dots + X_k = 1$ , kde  $k$  je počet faktorů a  $X_i$  jsou podíly jednotlivých složek směsi. Geometricky restrikce experimentálního prostoru odpovídá konvexnímu  $k - 1$  dimenzionálnímu polyhedronu<sup>5</sup>. Tyto plány lze označit za simplexové plány, které se využívají při modelování směsí. Otázka modelování směsí je velmi aktuální např. ve farmacii při řešení problematiky rozpustnosti účinných látek. Naprostá většina nových molekul je tvořena látkami nerozpustnými ve vodě a vodných roztocích, nicméně pro správný účinek je nutný transport vodným prostředím k cílovému místu (receptoru, buňce, orgánu). Ke zvýšení rozpustnosti látek je možné použít fyzikální metody (např. změna krystalinity), chemické metody (např. změna soli) nebo jejich kombinace. Ve většině případů se nejedná o klasickou formulaci ve formě přímého míchání nebo granulace, ale hovoříme o specializovaných systémech dodávání léků DDS<sup>6</sup>.

Velmi pěkné vysvětlení problematiky směsných experimentů lze nalézt v monografii [23] nebo [24]. Přístupný a praktický výklad této metodologie podává např. [3], nebo odbornější monografie [131].

Jedním z krucióálních témat je ve statistice bezesporu optimalizace. Existuje celá řada optimalizačních problémů, na které lze při řešení praktických problémů narazit. Jedním z příkladů může být minimalizace různých funkcí v případě různých metod shlukové analýzy [1], metoda SVM<sup>7</sup> [25] nebo i běžná maximalizace věrohodnosti<sup>8</sup> v případě hledání odhadů neznámých statistik.

Zde se však zaměříme na různé přístupy k řešení optimalizačního problému, který spočívá v identifikaci „optimálních operačních podmínek“, tedy oblasti nebo bodu v námi a priori vymezeném prostoru, při jehož dosažení získáme maximální, určitou konkrétní, nebo naopak minimální hodnotu response, případně „maximální“, určité konkrétní, nebo „minimální“ hodnoty jednotlivých elementů multiresponsního vektoru. Matematicky lze problém

<sup>5</sup>mnohostěnu, případně mluvíme i o simplexu

<sup>6</sup>DDS - Drug Delivery Systems

<sup>7</sup>Support Vector Machine

<sup>8</sup>Ve své podstatě i v průběhu řešení metody nejmenších čtverců lze úlohu formulovat prostřednictvím věrohodnostního přístupu a následně hledat optimum.

formulovat následovně:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^p} f(\mathbf{x}) &= i = 1, 2, \dots, M, \\ \text{vzhledem k } h_j(\mathbf{x}) = 0 &= j = 1, 2, \dots, J, \\ g_k(\mathbf{x}) \leq 0 &= k = 1, 2, \dots, K. \end{aligned}$$

kde jednotlivé elementy  $x_1, x_2, \dots, x_p$  vektoru  $\mathbf{x}$  jsou optimalizované proměnné, které mohou být obecně spojité, diskrétní nebo mohou být jejich kombinací. K řešení úlohy lze využít různých metod. Používají se například klasické numerické iterační algoritmy jako je modifikovaný Newtonův algoritmus nebo quasi-Newtonův algoritmus [146]. Pro svoji jednoduchost je často s výhodou používán nederivační Nelderův-Meadův simplexový algoritmus [101]. Dalším příkladem je v programovacím prostředí R dostupný algoritmus L-BFGS-L popsaný v práci [20] nebo metoda konjugovaných gradientů, popsaná spolu s kódem programu v monografii [100]. Pěkný přehled problematiky numerické optimalizace je podán v monografii [104]. Podobných numerických algoritmů existuje v současné době celá řada. Nevýhodou těchto algoritmů je však skutečnost, že není zaručeno nalezení globálního optima.

Dnes však existují, jak již bylo uvedeno výše, modernější algoritmy umožňující prohledání v podstatě celého uvažovaného parametrického prostoru a tím pádem i získání globálního optima a to s velkou pravděpodobností. Tyto algoritmy lze označit za heuristiky, které jsou velmi často inspirovány samotnou přírodou a zároveň jsou ovlivněny i náhodou, viz např. monografie [146].

Mezi takové přístupy patří například různé modifikace tzv. genetických algoritmů, simulované žíhání, tabu hledání, nebo optimalizace hejnem částic. Genetické algoritmy, o jejichž popularizaci se zasloužil především Holland [55], lze charakterizovat jako procedury, které jsou založeny na napodobení principu selekce, křížení a mutace. Výhodou stochastických přístupů, mezi které lze tyto heuristiky zařadit, je jejich schopnost nalézt globální optimum různě složitých funkcí, přičemž ty nemusí být ani diferencovatelné, ani spojité.

V případě RSM se lze setkat i s požadavkem na simultánní optimalizaci vektoru závislých proměnných. V takovém případě je možné využít postupu, který ve své práci navrhnul Derringer a Suich [30]. Tato procedura či její modifikace jsou dostupné například ve statistickém software jako je JMP, Design-Expert nebo v prostředí Matlab. Stejný přístup umožňuje i knihov-

na `desirability` [74] programovacího prostředí R.

Vhodnost výše zmiňované metodologie DOE a RSM a následné optimalizace je pro experimentální práci doložena v mnoha různých publikacích [11]. Příkladem může být práce [139], kde autoři využili metodologii DOE k optimalizaci specifity proby pro real-time PCR.

Další oblastí s širokou mírou využití může být komplexní problém související s optimalizací biotechnologických procesů, zejména pak v průběhu skreeningu produkčních kmenů, kdy se uplatní především částečné faktoriálové plány nebo následně při zvyšování produkce biologicky účinných látek, při kterých se využije metodologie RSM spolu s optimalizací kultivačních podmínek, viz např. práce [36]. Další uplatnění, a to zejména v případě směsných experimentů, lze spatřovat při formulaci a optimalizaci nových lékových forem.

## 2.2 Klasické přístupy k hodnocení experimentálních dat

### 2.2.1 Základní charakteristiky a příbuzné deskriptivní nástroje

Za klasické přístupy k hodnocení experimentálních dat lze označit popisné charakteristiky, které umožňují získat primární představu o poloze, variabilitě a distribuci sledovaných veličin, viz např. [4]. Ve své podstatě jde o přehledové statistiky, umožňující rychlou představu o situaci, přestože tyto statistiky mohou být značně zavádějící.

Mezi nejpoužívanější a zároveň asi nejzneužívanější se řadí prostý aritmetický průměr, směrodatná odchylka, minimum a maximum. To je do určité míry zapříčiněno dostupným komerčním statistickým softwarem, jakým je např. *Statistica*, *SPSS* nebo *NCSS*. Méně často bývají tyto doplněny o robustnější charakteristiky, jako jsou např. medián, modus, jednotlivé kvantily, IQR či např. Winsorizovaný průměr s určitým stupněm useknutí, voleným na základě tvaru rozdělení [144] a snižujícím ztrátu informace v porovnání s jednoduše ořezaným průměrem.

Právě robustnější charakteristiky mohou být adekvátnější při základním zpracování experimentálních dat, jenž se neřídí normálním rozdělením či sle-



dují např. asymetrické rozdělení s přítomností odlehlých hodnot.

Existují však i jiné „základní“ charakteristiky, které se využívají zejména pro popis dynamiky distribuce účinných látek, API<sup>9</sup>, v krvi organismu [60]. Tyto charakteristiky se zároveň ukázaly jako vhodné při popisu průběhu infekce vyvolané parazity, jak u hospodářských, tak laboratorních zvířat, viz práce [72].

Pokud jde o konfidenční intervaly ve tvaru  $\theta \pm \Delta$ , kde symbol  $\theta$  představuje odhad zkoumaného parametru a  $\Delta$  maximální přípustnou chybu, pak z celé plejády základních charakteristik bývají zpravidla využívány nejčastěji konfidenční intervaly pro střední hodnotu. Méně často se setkáváme s intervalovými odhady pro ostatní charakteristiky. V případně intervalových odhadů pro relativní četnosti, je zpravidla využíván asymptotický přístup ke konstrukci konfidenčního intervalu, přičemž jiné, méně časté a v některých případech adekvátnější přístupy lze nalézt například ve výborné monografii Agrestiho, viz [2]. Příkladem může být využití Willsonovo, Agresti-Coullovo nebo Clopperovo-Pearsonovo konfidenčního intervalu, namísto klasického Waldovo konfidenčního intervalu pro relativní četnosti, které se nezdá vyskytovat v observačně zaměřených publikacích. Tyto metody jsou dnes snadno dostupné prostřednictvím programovacího prostředí R, viz text [112], nebo monografie [10] a mohou být snadno využity.

### 2.2.2 Tradiční konfirmační metody

Mezi klasické přístupy k analýze experimentálních dat, lze dále řadit i konfirmační metody, resp. metody testování statistických hypotéz. Zde se prosazují především testy, umožňující komparaci mezi různými experimentálními skupinami, viz literatura [4] nebo [144]. Ve většině případů jde o parametrické testy, které lze vyjádřit prostřednictvím následujících nulových hypotéz:  $H_0 : \mu = \mu_0$ ;  $H_0 : \mu_{\text{kontrola}} = \mu_{\text{ošetření}}$  nebo např.  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ , kde  $k > 2$ , tedy pomocí jednovýběrového či různých variant dvouvýběrových  $t$ -testů, resp. celou další problematikou analýzy rozptylu. Analogicky pak pro relativní četnosti, případně pro regresní koeficienty či např. pro posouzení shody distribučních funkcí.

Zvláštním využitím výše zmíněného dvouvýběrového  $t$ -testu je tzv. TOST<sup>10</sup>

<sup>9</sup>Active Pharmaceutic Ingredients

<sup>10</sup>Two one side  $t$ -test

metoda [127], která našla uplatnění především při posouzení bioekvivalence, viz např. článek [107]. Zde je testovaná, resp. nulová, hypotéza koncipována jako  $H_0 : \mu_T - \mu_R \leq \theta_1 \cup \mu_T - \mu_r \geq \theta_2$ , přičemž její negací je tvrzení  $\theta_1 < \mu_1 - \mu_2 < \theta_2$ . Tedy, že se posuzovaný parametr, resp. rozdíl skutečných středních hodnot, pohybuje s předem danou spolehlivostí ve specifikovaných mezích. Právě popsany přístup se uplatňuje při posuzování generik, kdy o generickém preparátu mluvíme tehdy, jestliže se hodnoty testovaného přípravku pohybují v intervalu 80 - 125 % originálu. V případě léčiv s úzkým terapeutickým oknem je pro registraci generického léčiva interval zúžen na hodnoty 90-110%.

Testy výše zmíněných tradičních hypotéz jsou zpravidla parametrické a jsou vázány na dosti striktní předpoklady, které bývají v experimentální praxi nezdůvodněně porušeny. Za všechny lze uvést například problémy s normalitou dat či homoskedasticitou mezi jednotlivými výběry.

K těmto problémům se přidávají i specifická omezení, která jsou zejména v zootechnických oborech determinující. Příkladem za všechny může být např. délka reprodukčního cyklu u hospodářských zvířat, se kterou je spojeno obtížné získávání experimentálních dat a vede k malému počtu pozorování<sup>11</sup>.

Řešení těchto obtíží zpravidla v praxi spočívá ve využití neparametrických metod, umožňujících testování výše uvedených hypotéz. Sem se řadí zejména využití Mannova-Whitneova pořadového testu [144] a Kruskalova-Wallisova neparametrického testu [65]. Často neřešenou otázkou pak bývá problematika související s chybou druhého druhu, potažmo tedy se silou testu. Tato otázka do jisté míry souvisí i s problematikou rozsahu výběru(ů).

### 2.2.3 Klasický přístup k modelování závislosti

Dosti často se v praxi setkáváme se situací, ve které potřebujeme postihnout vztah mezi experimentálně naměřenými veličinami, mezi kterými odůvodněně předpokládáme existenci příčinné – kauzální vazby, přesněji statistické závislosti. V takovém případě zpravidla využíváme aparát regresní a korelační analýzy [28], přičemž se zpravidla snažíme modelovat  $E(y) = f(\mathbf{x}, \theta)$ .

<sup>11</sup>Obecně jsou malé soubory jednou z příčin nemožnosti aplikace některých moderních vícerozměrných statistických metod, neboť čelíme tzv. prokletí dimenzionality, viz [62].

Velmi často se autoři publikací uchylují k třídě lineárních modelů [28]. Asi nejjednodušším přístupem pak bývá využití prosté lineární regrese s chybovou složkou  $\epsilon \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ , která je asi nejpoužívanějším a zároveň nejzneužívanějším postupem.

Odhad regresních koeficientů je pak proveden zpravidla metodou nejmenších čtverců, tj. prostřednictvím minimalizace funkce  $\sum_i^n (y_i - \hat{y}_i)^2$ . Méně časté jsou v literatuře modernější, případně i robustnější, neparametrické přístupy, které budou zmíněny níže v textu.

Ne vždy si vystačíme s využitím lineárních modelů. Příčinou může být složitost modelovaného biologického systému nebo model musí splňovat určité omezení, případně je model výsledkem řešení diferenciální rovnice či diferenciálních rovnic. V takovém případě je možné využít nelineárních modelů [6], tj. modelů, které nejsou lineární kombinací parametrů<sup>12</sup>. Příkladem využití nelineárního modelu může být např. práce [136].

Někdy se můžeme v praxi setkat se situací, kde je závisle proměnnou veličina, která nevykazuje normální rozdělení. Například může sledovat poissonovo nebo binomické či negativně binomické rozdělení. V takovém případě lze za určitých podmínek použít zobecněné lineární modely. Pokud závisle proměnná vykazuje binomické rozdělení, pak lze použít např. logistický regresní model [53] umožňující predikovat binární závisle proměnnou. Právě takový model byl využit při modelování vlivu věku zvířat, technologie chovu a odstavu na výskyt kryptosporidiových infekcí v chovech prasat viz [102]. Blíže se lze o problematice GLM modelů dočíst např. v monografii [83] nebo [2], [53].

Jiným přístupem k modelování závislosti je aplikačně dosti rozšířená analýza přežití. Jejím hlavním cílem je modelování a analýza tzv. „time-to-event“ dat, blíže viz [89]. Asi nejjednodušším a v praxi nejužívanějším přístupem k modelování je v tomto případě tzv. Kaplanův-Meierův neparametrický

<sup>12</sup>Jako kritérium pro posouzení toho, zda jde o lineární či nelineární model, lze využít jednoduché pravidlo, které lze vyslovit následujícím způsobem. Pokud je parciální derivace regresní funkce podle některého z parametrů závislá na některém z parametrů regresního modelu, pak je model nelineární. Například regresní model  $y = \beta_1 + \beta_2[1 - \exp(\beta_3 x)]$  je nelineární, neboť je zřejmé, že:

$$\frac{\partial y}{\partial \beta_1} = 1 \quad \frac{\partial y}{\partial \beta_2} = 1 - \exp(\beta_3 x) \quad \frac{\partial y}{\partial \beta_3} = -\beta_2 x \exp(\beta_3 x).$$

Model je tedy nelineárním.

přístup<sup>13</sup>, který k odhadu funkce přežití využívá následující model:

$$\hat{S}(t) = \prod_{t_i \leq t} (1 - \hat{q}_i) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right),$$

kde  $n_i$  je počet dosud žijících v čase  $t_i$  a  $d_i$  je počet jedinců zemřelých v témže čase. Pomocí tohoto přístupu lze např. analyzovat přežívání skupin v různých podmínkách a následně i testovat odlišnost v jejich přežívání s využitím tzv. Log-rank testu [50]. Tento přístup se uplatnil například v práci [68].

#### 2.2.4 Klasické metody vícerozměrné analýzy dat

Velmi často v průběhu experimentu nesledujeme pouze osamocenou veličinu, ale spíše množinu proměnných, přičemž nás zajímají jejich vzájemné statistické závislosti, resp. korelační vazby. V takovém případě využíváme metody, které lze souhrnně označit za vícerozměrné. Ty jsou ve své podstatě charakterizovány simultánním nahlížením na soubor proměnných. Rovněž i každé pozorování nyní náleží do prostoru s vyšší dimenzí, přičemž se ne-žádka jedná o  $\mathbb{R}_p$ , kde  $p \geq 2$ , jinak řečeno, to co byl skalár, je nyní vektor. Z hlediska výpočetní statistiky se zde setkáváme s masivním využitím maticové algebry a nasazením výpočetní techniky. V některých případech se dnes řeší i optimalizace numerických výpočtů prostřednictvím paralelizace potřebných numerických výpočtů a s tím spojeným dělením jader na výpočetních stanicích. Ale věnujme se nyní začátkům, které sahají až do první třetiny dvacátého století<sup>14</sup>.

Za tradiční vícerozměrné metody zpravidla označujeme metody redukce dimenzionality, kam řadíme především analýzu hlavních komponent<sup>15</sup> velmi dobře popsanou v monografii [59], metody vícerozměrného škálování [12], faktorovou analýzu [5] a v neposlední řadě i francouzskou školou propagovanou korespondenční analýzu, viz monografie [47], [48]. Pokud jde o PCA, jejím cílem je prostřednictvím výrazně menšího počtu latentních, nikoliv přímo pozorovaných proměnných, které jsou lineární kombinací původních proměnných, vyjádřit co možná nejvíce variability obsažené v datech. Použití PCA především spočívá ve vhodné reprezentaci vícerozměrných dat do

<sup>13</sup>Hovoříme o Kaplanovo-Meierovo odhadu, nebo také „součinovém limitním odhadu“. Existují však i jiné modely, jako je například model Coxové

<sup>14</sup>Paradoxně jedna z nejpoužívanějších exploračních metod, totiž PCA je datována zhruba do tohoto období. Její počátky lze nalézt v psychologických aplikacích

<sup>15</sup>PCA - Principal component analysis

prostoru s výrazně nižší dimenzí, která umožňuje i jejich grafické znázornění. Podobný účel splňují i další výše zmiňované metody (MDS, CA).

Tradiční a dosti širokou skupinou v praxi využívaných vícerozměrných metod jsou metody, resp. algoritmy shlukové analýzy, viz [61] nebo např. [1]. Tyto metody nalézají široké uplatnění nejenom v genomice, chemometrii a taxonomii, kde jde především o skupinu hierarchických shlukovacích metod zmiňovaných, např. v literatuře [121], a to zejména pro jejich explicitnost vyjádřenou prostřednictvím dendrogramů zachycujících průběh shlukování. Je třeba si uvědomit, že výsledek shlukování je odvislý jak od použitého shlukovacího algoritmu, tak i použité metricky, reprezentující nepodobnosti shlukovaných objektů. Na tomto místě je třeba také zdůraznit skutečnost, že pro různé praktické problémy je třeba adekvátně volit použitou metriku! Ne vždy je vhodná např. prostá euklidovská metrika. Zejména u moderních genomických metod je tato otázka dosti důležitá, nicméně tento problém je stále předmětem zkoumání a není dosud uspokojivě vyřešen. Rovněž složitou otázkou je volba správného počtu shluků<sup>16</sup>.

Mezi další vícerozměrné metody lze zařadit i skupinu metod souhrnně označované za klasifikační metody. V podstatě lze říci, že tak jako analýza rozptylu, tak i klasifikační metody, umožňují modelování vztahu mezi skupinou kvantitativních proměnných<sup>17</sup> na jedné straně a kvalitativní proměnnou na straně druhé. Podstatný rozdíl však spočívá ve směru studované závislosti. Zatímco u analýzy rozptylu je závisle proměnnou kvantitativní proměnná či skupina kvantitativních proměnných, pak u klasifikačních metod je tomu právě naopak. Zde představuje kvalitativní proměnná závisle proměnnou, která může vykazovat obecně  $g$  různých tříd - populací, které lze označit prostřednictvím symbolů  $\Pi_1, \Pi_2, \dots, \Pi_g$ , kde  $g \geq 2$ .

Cíl klasifikačních metod pak spočívá v odvození klasifikačních pravidel umožňujících přiřazení objektu  $\mathbf{x}_i$ , u kterého byly pozorováním naměřeny hodnoty  $x_{i1}, x_{i2}, \dots, x_{ip}$ , do některé z  $g$  populací. Pochopitelně se při klasifikaci snažíme co nejméně chybovat.

Klasifikační pravidla ve své podstatě představují rozdělení prostoru  $\mathcal{X}$  do

---

<sup>16</sup>Metody shlukové analýzy jsou považovány za tzv. „metody bez učitele“, kde není a priori znám počet shluků. Oproti tomu, např. klasifikační metody, původně označované za tzv. „Pattern Recognition“ metody či někdy i ne zcela správně a obsírně za „Data-mining metody“, touto informací zpravidla disponují, viz tréninková množina dat, vs. testovací, případně validační množina dat, jejichž sestavení v podstatě determinuje odvození výsledných klasifikačních pravidel.

<sup>17</sup>Případně mixem kvantitativních a kvalitativních proměnných

$g$  subprostorů  $\mathcal{X}_j$ ,  $j = 1, 2, \dots, g$ , přičemž zpravidla platí, že  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ , pro  $i \neq j$ . Klasifikační pravidlo můžeme vyjádřit následujícím způsobem: „Zařaď  $\mathbf{x}$  do  $\Pi_j$  pokud  $\mathbf{x} \in \mathcal{X}_j$ , pro  $j = 1, 2, \dots, g$ .“ Intuitivně můžeme předpokládat, že čím větší část hustoty pravděpodobnosti se bude pro populaci  $\Pi_j$  nacházet v subprostoru  $\mathcal{X}_j$ , tím kvalitněji bude prováděná diskriminace. Existuje mnoho přístupů k řešení výše nastíněného problému. Zajímavé přístupy ukazuje např. Duda a kol. ve své monografii [31].

Za dnes již klasické metody, řešící klasifikační problém lze označit lineární diskriminační analýzu<sup>18</sup> (LDA) popsanou např. v [92] nebo její kvadratickou formu popsanou tamtéž. LDA vychází ze znalosti hustot pravděpodobnosti jednotlivých skupin  $\Pi_i$ , přičemž využívá Bayesovo pravidlo [80]. Označíme-li symbolem  $f_k(\mathbf{x})$  podmíněnou hustotu pravděpodobnosti v  $k$ -té populaci a symbolem  $\pi_k$  apriorní pravděpodobnost „výskytu“  $k$ -té skupiny - populace, pak je zřejmé, že podmíněná pravděpodobnost

$$P(\Pi_k|\mathbf{x}) = f_k(\mathbf{x})\pi_k / \sum_{i=1}^g f_i(\mathbf{x})\pi_i \propto f_k(\mathbf{x})\pi_k .$$

V případě, že pro kovarianční matice platí  $\Sigma_i = \Sigma$ , pro  $\forall i$ , a jednotlivé hustoty  $f_i(\mathbf{x})$ , pro  $\forall i$ , mají  $N_p(\boldsymbol{\mu}_i; \Sigma)$ , stačí pro klasifikaci mezi dvěma skupinami stanovit pouze podíl,  $P(\Pi_k|\mathbf{x})/P(\Pi_l|\mathbf{x})$ . Obdržíme tím pro jednotlivé skupiny tzv. lineární diskriminační funkce, kde ty mají tvar:

$$\delta_k(\mathbf{x}) = \mathbf{x}^t \Sigma^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^t \Sigma^{-1} \boldsymbol{\mu}_k + \log \pi_k .$$

Objekt  $\mathbf{x}$  je následně zařazen do skupiny  $k$ , pokud platí, že

$$\Pi_k = \arg \max_i \delta_i(\mathbf{x}) .$$

V případě, že jsou kovarianční matice jednotlivých skupin rozdílné, je adekvátnější použít kvadratickou diskriminační analýzu<sup>19</sup> [92], případně neparametrický Fisherův diskriminační přístup [40], jehož výhodou je skutečnost, že není nutná znalost parametrických forem rozdělení u jednotlivých populací  $\Pi_1, \Pi_2, \dots, \Pi_g$ . Fisherův přístup je založen na hledání nadrovin, které z jistého pohledu co nejlépe separují  $g$  libovolných skupin - populací.

---

<sup>18</sup>LDA

<sup>19</sup>QDA

Ve své podstatě jde o nalezení lineární kombinace jednotlivých proměnných, tj.  $\mathbf{a}^t \vec{x}$ , umožňujících maximalizovat podíl meziskupinového a vnitroskupinového rozptylu. Řešením jsou složky vektoru  $\mathbf{a}^t = [a_1, a_2, \dots, a_p]$ , maximalizující podíl<sup>20</sup> tzv. Rayleighova kvocientu:  $\Gamma = \mathbf{a}^t \mathbf{B} \mathbf{a} / \mathbf{a}^t \mathbf{W} \mathbf{a}$  při omezující podmínce  $\mathbf{a}^t \mathbf{W} \mathbf{a} = \mathbf{1}$ . Výše zmíněné klasifikační metody jsou popsány v celé řadě publikací, viz [92] nebo aplikačně zaměřené literatuře [137]. Výhodou těchto metod je jejich široká dostupnost prostřednictvím statistického softwaru jako je R, SAS, Statistica, NCSS nebo SPSS.

Za méně známé klasifikační přístupy lze považovat například tzv. ZeroR klasifikátor, OneR klasifikátor [103] či Naivní bayesovský klasifikátor [58]. Jejich využití se zpravidla omezuje na posouzení výkonnosti jednotlivých klasifikátorů.

Koncepčně jiným přístupem je metodologie klasifikačních a regresních stromů, tzv. CART<sup>21</sup>, využívající rekurzivní dělení prostoru, viz práce Morgana a Sonquista [91]. Na základě jejich práce se vyvinuly dva velmi populární algoritmy CART a C4.5. První ze jmenovaných byl navržen Breimanem a kol. [17], druhý pak nezávisle Quinlanem [56]. Společnou charakteristikou obou navržených přístupů je, jak již bylo řečeno výše, provedení disjunktního rozdělení prostoru  $\mathcal{X}$  do množiny, obecně,  $p$ -rozměrných rovnoběžnostěnů. Ty si lze představit jako subprostory s následným přiřazením určité třídy každému z  $p$ -dimenzionálních rovnoběžnostěnů. Pro rozklady prostoru  $\mathcal{X}$  bude platit  $\cup_{j=1}^g \mathcal{X}_j = \mathcal{X}$  a současně  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ , pro  $\forall i, j : 1 \leq i, j \leq g$  a  $i \neq j$ . Vznikají tak klasifikační pravidla, která lze s výhodou znázornit prostřednictvím klasifikačního stromu<sup>22</sup>. Cíl však nelze spatřovat v pouhém získání přesného klasifikačního pravidla, ale také v získání informací o vlivu a vztazích jednotlivých proměnných, neboť ty determinují charakter klasifikačních pravidel. Výsledný klasifikační strom je třeba hodnotit podle tří kritérií. Jsou jimi především predikční přesnost, velikost stromu a jeho srozumitelnost. Jednotlivá kritéria lze do určité míry ovlivnit při konstrukci klasifikačního stromu vhodnou volbou hyperparametrů, blíže viz [17]. Výhodou klasifikačních stromů je především to, že nejsou vázány

<sup>20</sup>Tento podíl je také znám pod názvem Fisherovo diskriminační kritérium.

<sup>21</sup>CART - Classification and Regression Trees

<sup>22</sup>Klasifikační strom je tedy acyklický graf tvořený uzly. V souvislosti s popisem klasifikačního stromu se vžily pro různé uzly ve stromu specifické názvy. Jde především o pojmy kořen, uzel a list. Kořen stromu je tvořen uzlem obsahujícím všechny objekty. Uzlem se pak rozumí každá hladina, na níž je provedeno binární štěpení. Listem pak označujeme každý, již dále neštěpený, tj. terminální uzel.

na splnění předpokladů o tvaru rozdělení. Další výhodou je i možnost implicitního vyjádření interakcí plynoucí ze směny prediktorů, vyskytujících se v klasifikačních pravidlech na jednotlivých úrovních štěpení stromu. Výhodou je i schopnost pracovat s neúplnými daty. Naopak, určitou nevýhodou klasifikačních stromů je však jejich nestabilita. Ta je odstraňována prostřednictvím modernějších technik, viz kapitola níže.

### 2.3 Některé moderní přístupy k hodnocení experimentálních dat

Při základním zpracování dat často dochází i k nevyužití grafického aparátu, který může mnohdy značně přispět k pochopení distribuce dat, tj. k pochopení toho, jakých hodnot s jakými pravděpodobnostmi může sledovaná veličina nabývat. Zde lze především upozornit na možnosti použití odhadů hustot pravděpodobnosti získaných prostřednictvím jádrových odhadů [128]. Rovněž dochází i k obecně menšímu využití grafického znázornění distribuce dat pomocí empirických distribučních funkcí, přičemž obě zmiňované funkce  $\hat{F}_X(x)$  i  $\hat{f}_X(x)$  v sobě obsahují v podstatě veškerou informaci obsaženou v datech. V případě explicitnějších a názornějších hustot může být problémem volba šířky vyhlazovacího okna, viz literatura [128], tak i dalších hyperparametrů, kterým může být typ vyhlazovacího jádra či stupeň vyhlazení.

Jak již bylo řečeno, při řešení některých praktických úkolů plynoucích z experimentální činnosti může být problémem menší počet experimentálních subjektů, případně problémy s normální distribucí analyzovaných dat. V takovém případě lze použít modernější konfirmační přístupy, které jsou založeny na permutačních testech, blíže viz [34] nebo v monografii [26]. V případě problematického využití jednofaktorové analýzy rozptylu, ať již ANOVy<sup>23</sup> či MANOVy<sup>24</sup>, lze s výhodou využít např. metodu DISCO<sup>25</sup> navrženou Rizzem a Székelym [116], která je založena na měření celkové disperse dat. Metodu můžeme chápat jako zobecnění klasické analýzy rozptylu, přičemž při nastavení tzv. „power“ parametru  $\alpha$  na hodnotu 2 získáme stejné řešení jako v případě klasické analýzy rozptylu. Metoda DISCO má i tu výhodu,

---

<sup>23</sup>Analysis of Variance

<sup>24</sup>Multivariate Analysis of Variance

<sup>25</sup>DISCO - Distance Components



že je implementována i v programovacím prostředí R.

Pokud jde o modelování vztahů mezi proměnnými prostřednictvím regresních modelů, lze říci, že je v současnosti dosažitelná a v praxi uplatnitelná celá plejáda zajímavých přístupů.

Jedním z nich může být např. kvantilová regrese zavedená Koenkerem [64], ve které jsou  $\tau$ -procentní kvantilové regresní koeficienty v případě lineárního kvantilového modelu  $Q_Y(\tau|\mathbf{X} = \mathbf{x}) = \mathbf{x}^t \boldsymbol{\beta}_\tau$  získány následovně:

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^t \boldsymbol{\beta}) \right\},$$

kde funkce  $\rho_\tau(r)$  je definována pomocí kvantilové ztrátové funkce  $\rho_\tau(r) = r(\tau - \mathbf{I}(r < 0))$ . Metoda je v současnosti rovněž dostupná prostřednictvím různého statistického software.

Dalším moderním regresním nástrojem je například tzv. LASSO<sup>26</sup> [57], jehož pomocí lze překonávat problém multikolinearity. V průběhu této techniky je prováděna L1 regularizace přidávající penalizační parametr využívající absolutních hodnot odhadovaných regresních parametrů, jinými slovy je minimalizována funkce:

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 - \lambda \sum_{j=1}^p |\beta_j|.$$

Oproti hřebenové regresi tak LASSO umožňuje v podstatě „automatický“ výběr proměnných do regresního modelu.

Zcela jiná a dnes dynamicky se rozvíjející oblast analýzy dat je funkcionální přístup k analýze dat<sup>27</sup>. Tato metodologie se od tradičního pojetí odlišuje hlavně tím, že data jsou zde chápána jako funkce, nikoliv jako body či vektory, viz tradiční pojetí. Základní referencí mohou být monografie [113], [37] nebo aplikačně založená [114]. V současné době lze využít funkcionální přístup k řešení různých problémů, jakými je např. analýza dynamiky růstu<sup>28</sup> nebo změna senzorických vlastností v závislosti na čase, jak

<sup>26</sup>Least Absolute Shrinkage and Selection Operator

<sup>27</sup>FDA - Functional Data Analysis

<sup>28</sup>Příčemž lze na základě odhadnutých trajektorií růstu zkonstruovat i průběhy rychlosti či zrychlení prostřednictvím derivací funkcí, za předpokladu, že je modelovaná funkce alespoň dvakrát derivovatelná, což úzce souvisí s nastavením parametrů při modelování dat, viz spiny a jejich řád a počet uzlů. Blíže viz [113], [37]

vysvětluje článek [8].

Jiným problémem může být např. regresní analýza, v rámci které lze rozeznávat koncepčně tři kvalitativně různé kategorie modelů, dle toho jak nahlížíme na jednotlivé složky regresního modelu. Tyto kategorie lze popsat jako „Scalar-on-function regression“

$$y_i = \int \beta(s)x_i(s)ds + \epsilon_i ,$$

zde jsou regresory funkce/křivky a závisle proměnná je skalár. Odhadnutým parametrem je funkce  $\beta$ . Dalším je typ „Function-on-scalar regression“

$$y_i(t) = \sum_{k=1}^p x_{ik}\beta_k(t) + \epsilon_i(t)$$

a třetí, tzv. „The Function-on-function regression“

$$y_i(t) = \int \beta(t,s)x_i(s)ds + \epsilon_i(t) ,$$

blíže viz monografie [73]. Celá metodologie funkcionální analýzy se v současnosti dosti rozvíjí.

Další výše diskutovanou oblastí byla vícerozměrná analýza dat. V této souvislosti lze říci, že v současnosti dochází k výraznému nárůstu sledovaných a měřených dat, a to zejména v oblastech jako jsou chemometrie, genomika, metabolomika, či proteomika. Obecně vzrůstá potřeba správného hodnocení velkého množství, mnohdy do velké míry korelovaných proměnných, s potřebou jejich následné vizualizace. Ta se zpravidla děje prostřednictvím klasické či modifikované analýzy hlavních komponent. Takovým příkladem může být pravděpodobnostní PCA, viz [133] nebo [105]. Ve velké míře se prosazují např. OPLS<sup>29</sup> metody [134] nebo [9]. Dále metody shlukové analýzy, PLS<sup>30</sup> regrese nebo i metody založené na grafech a sítích jako např. GGM<sup>31</sup> k popisu vzájemných závislostí v metabolických drahách, jak uvádí například článek [7]. Rovněž vzrůstá potřeba vhodné vizualizace. Zde se uplatňují „heatmapy“, viz [63] nebo např. v metabolomice „volcano“ grafy [75].

---

<sup>29</sup>OPLS - Orthogonal Projections to Latent Structures

<sup>30</sup>Partial Least Squares

<sup>31</sup>GGM - Gaussian graphical models

Pokud jde o klasifikaci, pak lze za modernější přístupy považovat celou řadu metod lišících se principem a předpoklady, ze kterých tyto metody vycházejí. Za všechny lze jmenovat například tzv. regularizovanou diskriminační analýzu<sup>32</sup>(RDA) navrženou Jerome Friedmanem [44]. Jeho metoda je založená na kompromisním přístupu, který využívá konvexní kombinace kovariančních matic. Řešení lze chápat jako vhodnou kombinaci výše zmíněného přístupu pomocí metod LDA a QDA. V případě RDA je regularizace skupinové kovarianční matice  $\Sigma(\lambda; \gamma)_k$ , resp. jejího odhadu  $\hat{\Sigma}(\lambda; \gamma)_k$ , provedena prostřednictvím předpisu

$$\hat{\Sigma}_k(\lambda; \gamma) = (1 - \gamma)\hat{\Sigma}_k(\lambda) + (\gamma/p)\text{stopa}[\hat{\Sigma}_k(\lambda)]\mathbf{I} ,$$

kde symbol  $\hat{\Sigma}_k(\lambda) = (1 - \lambda)\hat{\Sigma}_k + \lambda\hat{\Sigma}$ . Regularizační parametr  $\lambda$ , kde  $0 \leq \lambda \leq 1$ , pak ovlivňuje stupeň deformace<sup>33</sup> skupinové kovarianční matice  $\hat{\Sigma}_k$  „směrem“ ke společné kovarianční matici  $\hat{\Sigma}$ . Takto získané kovarianční matice jsou pak uplatněny v průběhu klasifikace. Při jejich aplikaci získáváme diskriminační funkce, které mají následující tvar

$$\delta_k(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_k)^t \hat{\Sigma}_k^{-1}(\lambda; \gamma)(\mathbf{x} - \boldsymbol{\mu}_k) + \ln |\hat{\Sigma}_k(\lambda; \gamma)| - 2 \ln \pi_k$$

pro  $k = 1, 2, \dots, g$ . Parametry  $\lambda$  a  $\gamma$  jsou voleny simultánně tak, aby bylo dosaženo co možná nejmenší nevychýlené chyby vzniklé nesprávnou klasifikací. Postupy pro volbu regularizačních parametrů jsou založeny na cross-validaci a bootstrapu, blíže viz literatura [33] a [44].

Jiným moderním přístupem je využití jádrových odhadů hustot při současné volbě vyhlazovací matice, viz [32]. Princip lze stručně formulovat následovně. Předpokládáme, že opět máme k dispozici  $g$  populací  $\Pi_1, \Pi_2, \dots, \Pi_g$ , které lze popsat pomocí hustot  $f_k$ ,  $k = 1, \dots, g$  s apriorními pravděpodobnostmi  $\pi_k$ ,  $k = 1, \dots, g$ . Zařazení objektu  $\mathbf{x}$  z výběrového prostoru provedeme na základě Bayesova diskriminačního pravidla, tedy:

$$\text{Zařad } \mathbf{x} \text{ do populace } \Pi_i, \text{ pokud } \Pi_i = \underset{k \in 1, 2, \dots, g}{\operatorname{argmax}} \pi_k f_k(\mathbf{x})$$

s tím, že v případě jádrově založené diskriminační analýzy je příslušná hustota  $f_k$  odhadována prostřednictvím jádrového odhadu hustoty  $\hat{f}_K(\mathbf{x}, \mathbf{H}_k)$ :

$$\hat{f}_k(\mathbf{x}, \mathbf{H}_k) = n_k^{-1} \sum_{i=1}^{n_k} K_{\mathbf{H}_k}(\mathbf{x} - \mathbf{x}_{ki}) ,$$

<sup>32</sup>RDA - Regularized discriminant analysis

<sup>33</sup>redukce, smrštění

kde  $K(\mathbf{x})$  je jádro, které je symetrickou hustotou pravděpodobnosti. Vyhla-  
zovací matice  $\mathbf{H}$  je pozitivně definitní maticí typu  $p \times p$ , obsahující šířky  
vyhlazovacích oken<sup>34</sup> pro jednotlivé prostory, jejichž volba ovlivňuje výsled-  
nou kvalitu diskriminace a tvar vyhlazovacího jádra, jak se uvádí v [142].  
Jedná se tak o ladící parametr metody. Vyhla-zovací matici  $\mathbf{H}$  volíme tak,  
aby bylo dosaženo co možná nejmenší střední čtvercové integrované chyby  
 $MISE$ , definované jako

$$MISE(\mathbf{H}) = \mathbb{E} \left( \int_{\mathbb{R}^p} [\hat{f}(\mathbf{x}; \mathbf{H}) - f(\mathbf{x})]^2 d\mathbf{x} \right).$$

Optimální volbou je pak tedy  $\mathbf{H}_{MISE} = \underset{\mathbf{H}}{\operatorname{argmax}} MISE(\mathbf{H})$ . Jádro  $K_{\mathbf{H}}(\mathbf{x})$   
stanovíme jako  $|\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2} \mathbf{x})$ . Jak uvádí Duong [32], volba jádra není  
až tak klíčová.

Častou volbou je např. Gaussovské jádro  $K(\mathbf{x}) = (2\pi)^{-p/2} \exp(-\frac{1}{2} \mathbf{x}^t \mathbf{x})$ ,  
viz např. [142]. Pro úplnost apriorní pravděpodobnost  $\pi_k$  je ve výše uve-  
dené Bayesovském pravidle odhadována prostřednictvím podílu  $n_k/n$ , kde  
 $n = \sum_{k=1}^g n_k$ . Výhodou je implementace výše uvedených metod v knihov-  
nách MASS [137], mda [52] a ks programovacího prostředí R [111].

V případě neparametrických klasifikačních metod byla odstraněna nestabi-  
lita stromových klasifikátorů prostřednictvím modernějších technik za-  
ložených na souborném hlasování, viz metodologie „Random Forest“ (RF)  
navržená Briemanem [19]. Metodologii náhodných lesů lze popsat jako kom-  
binaci stromových prediktorů takových, že každý strom je vybudován na zá-  
kladě náhodně a nezávisle vybraných dat se stejným rozdělením pro všechny  
stromy v lese. Klasifikace je pak umožněna souborným hlasováním na zá-  
kladě „hlasů“ jednotlivých stromů v lese. Tento přístup lze využít i pro  
regresní úlohy. Výhodou je odstranění nestability, kterou trpí jednotlivé  
stromy, viz metoda CART [17]. Další výhodou RF je, že umožňuje odhadovat  
významnosti jednotlivých proměnných použitých v odvození klasifikačních  
či regresních pravidel. Jiným přístupem jsou pak metody nazývané Bagging  
[18] a Boosting [43].

Existuje celá řada dalších klasifikačních algoritmů, např. diskriminační ana-  
lýza založená na analýze modů či modelově založená diskriminační analýza  
[41]. Zajímavé jsou i přístupy k diskriminaci založené na jádrových odha-  
dech, o kterých se můžeme dočíst například v monografii [51]. Zcela jiným,

---

<sup>34</sup>tzv. bandwidth matrix

a neméně zajímavým nepravděpodobnostním přístupem je metoda SVM<sup>35</sup>, která při řešení klasifikačního problému využívá optimalizační přístup. Koncepte SVM je založena na hledání separujících nadrovin, blíže viz [25]. Nezájemné nejsou ani v chemometrii využívané PLS a OPLS metody, konkrétněji PLS-DA či asi z hlediska efektivnosti diskriminace výkonnější OPLS-DA, viz článek [9], která se široce uplatňuje v chemometrii, metabolických či transkriptomických studiích. Zajímavé je i využití klasifikačních metod<sup>36</sup> při modelování vztahů mezi strukturou a biologickou aktivitou látek v tzv. QSAR<sup>37</sup> studiích, jak se uvádí v přehledovém článku [130].

## 2.4 Další směřování ve využití statistických metod v experimentální praxi

Lze říci, že na základě získaných znalostí o problematice kvantitativního hodnocení experimentálních dat, lze v diskutované oblasti předpokládat přibližně následující vývoj ve využívání statistických metod.

Vzhledem k dostupnosti různých statistických či data-miningových metod se s velkou „pravděpodobností“ začnou ve větší míře prosazovat modernější přístupy k hodnocení dat. Nicméně půjde především o metody, které budou úzce navázány na softwarově dostupné nástroje, přičemž se dá předpokládat značný příklon odborné komunity k využití freewarového programovacího prostředí R a software Python, viz například Bioconductor, či v případě Pythonu k distribuci s názvem Anaconda, jejíž součástí je i modifikace nazvaná Biopython, a to především pro jejich dostupnost, versatilitu, relativní jednoduchost skriptovacího jazyka a freewarový charakter obou platform. Tento trend bude zvláště výrazný u „-omic“ disciplín, ve kterých není problém v získání velkého počtu experimentálních dat ( $n \ll p$ ).

Při standardním hodnocení fenotypových parametrů v zootechnice však bude nadále limitujícím faktorem především počet dostupných zvířat, která budou k dispozici v observačních studiích, či pro případnou experimentální činnost. Omezením bude i případná časová náročnost experimentů. Za další omezení lze označit finanční náročnost či případné etické otázky. Vzhledem

---

<sup>35</sup>SVM - Support Vector Machine

<sup>36</sup>Zde se již zpravidla mluví o data-miningových metodách.

<sup>37</sup>QSAR - Quantitative Structure-Activity Relationship

k těmto limitujícím faktorům lze předpokládat větší využití modernějších, neparametrických a robustnějších metod, nicméně zde budou hrát prim stále klasické statistické metody jako je analýza rozptylu, regresní analýza či lineární smíšené modely a podobně.

## Komentář k publikovaným pracím

Níže je uveden autorský komentář k publikovaným pracím. Odkazy, resp. citace, odpovídají zařazení v použité literatuře, která je uvedena na konci práce. Odkazy na práce u nichž byl autor předložené práce spoluautorem jsou vyvedeny tučně.

V níže komentovaných pracích (komentář ukončený symbolem ■) [135], [68] a [120] byl využit zejména klasický statistický aparát, případně metodologie uplatnitelná pro bioekvivalenční studie, viz publikace [84], [85], [86] (tyto tři publikace zde nejsou vzhledem k zaměření habilitační práce rozebírány) nebo níže diskutovaná práce [107].

Vliv různého druhu krmiva na obsah tuku u tříletého kapra (*Cyprinus carpio* L.) byl sledován v publikaci [135]. Výsledky plynoucí z měření pomocí fatmetru ukázaly, že se obsah tuku různí v průběhu růstové sezóny při použití různých obilovin, přestože jednotlivé dávky byly upraveny tak, aby měly stejnou stravitelnou energii. Nejvíce podkožního tuku bylo zjištěno v případě dokrmování kukuřicí. Naopak, nejmenší podíl podkožního tuku byl zjištěn v případě kontrolní skupiny. Skupina dokrmovaná žitem vykazovala vyšší obsah tuku než skupina přikrmovaná prostřednictvím tritikále.

Při vyhodnocení byly využity klasické metody statistického zpracování dat. Konkrétněji byla využita základní popisná statistika spolu s jednofaktorovou analýzou rozptylu. Homoskedasticita byla ověřována prostřednictvím Levenova testu. Pro identifikaci signifikantně různých dvojic krmiv byl využit tzv. Tukeyovo HSD test na hladině významnosti  $\alpha = 0,05$ .

Z hlediska statistické metodologie a interpretace vlastních výsledků by mohlo být dnes zajímavé využít moderní aparát analýzy funkcionálních dat [37]

spolu s možností vlastní analýzy derivací prvního a druhého řádu funkcí skupinových průměrů, případně i posouzení rozdílností prostřednictvím funkcionální ANOVy, či funkcionální ANOVy popsané v podané práci autora, viz [95], která je implementována v knihovně GET programovacího prostředí R na tzv. CRANu<sup>1</sup>. ■

V rámci publikace [68] byl studován vliv aktivovaných CD4+ T a CD8+ T lymfocytů na průběh infekce vyvolané žaludeční kryptosporidií *Cryptosporidium proliferans* (v té době známé jako *C. muris* kmen TS03) u imunokompetentních a imunodeficitních laboratorních myší (*Mus musculus*). Pro účely zjištění vlivu CD8+ T lymfocytů na schopnost hostitele vyrovnat se s infekcí byl sledován průběh infekce zahrnující intenzitu infekce. Ta byla charakterizována prostřednictvím plochy pod křivkou (AUC) stanovenou pomocí klasického trapezoidálního pravidla. Jako další charakteristiky dynamiky infekce byl sledován čas spolu s hodnotou maximální intenzity infekce, tj.  $t_{max}$  a  $C_{max}$ .

Vzhledem k značné finanční náročnosti prováděných experimentů a malému počtu laboratorních myší, spolu s problematickou normalitou získaných dat, byl využit neparametrický Kruskalův–Wallisův test. Pro vícenásobné párové srovnávání byl použit Mannův–Whitneyův test s následnou Bonferroniho korekcí, s cílem dodržet deklarovanou hladinu významnosti.

Pro posouzení přežívání v jednotlivých experimentálních skupinách byla použita analýza přežití, konkrétněji byl použit Kaplanův–Meierův neparametrický přístup. K testování rozdílnosti neparametricky odhadnutých funkcí přežití byl následně použit tzv. log-rank test, přičemž parametr  $\rho$  byl nastaven na nulovou hodnotu. Výsledky provedených analýz prokázaly, že zatímco imunodeficitní hostitelé, kteří byli rekonstituováni aktivovanými a neaktivovanými CD4+ nebo aktivovanými CD8+ T lymfocyty se vyléčili z infekce, rekonstituce neaktivovanými CD8+ T lymfocyty neměla žádný vliv na průběh onemocnění. Rekonstituce CD4+ T buňkami vedla k poklesu množství vylučovaných oocyst a ke zkrácení patentní periody. V neposlední řadě tato práce ukázala na významný vliv aktivace CD8+ T lymfocytů v ochraně hostitele proti žaludeční kryptosporidiióze.

Výše popsany statistický přístup při charakterizaci intenzity infekce byl použit i v dalších experimentech sledujících průběh infekce *Cryptosporidium avium* u různých ptačích hostitelů [54], *C. tyzzeri* u myší domácích [70] nebo při ověření věkové specifiity *C. scrofarum* u prasat, viz [72]. Náš přístup

---

<sup>1</sup>The Comprehensive R Archive Network



umožňuje charakterizovat interakci mezi parazitem a hostitelem na základě průběhu infekce, která je ovlivněna celou řadou faktorů, jako je druh hostitele, věk, pohlaví či stavem imunitního systému. V práci [54] bylo prokázáno, že zatímco věk u kura domácího (*Gallus gallus* f. *domesticus*) ovlivňuje průběh infekce vyvolané *C. avium*, u kachen (*Anas platyrhynchos* f. *domesticus*) tento vliv nebyl zaznamenán. Z výsledků publikace (myši) vyplývá, že průběh infekce (charakterizovaný prostřednictvím  $AUC$ ,  $t_{max}$  a  $C_{max}$ ) vyvolaný izoláty *C. tyzzeri* pocházejících z myši východoevropských (*Mus musculus musculus*) a myši západoevropských (*Mus musculus domesticus*) se u jednotlivých testovaných druhů hostitelů neliší (*M. m. musculus* a *M. m. domesticus*), přestože v přírodě nedochází k přirozené infekci myši západoevropských izoláty kryptosporidií z myši východoevropských a obráceně. ■

Článek [120] se věnuje problematice ztrát rybích společenstev způsobených povodňovými škodami na rybníčních akvakulturách v letech 2002, 2006, 2009 a 2013 na území České Republiky. Celkové škody (ztráty) rybích společenstev byly za tyto období odhadnuty, na základě údajů z rybářských podniků, přibližně na 54,2 % celkové populace. Ukázalo se, že kapr je mnohem méně náchylný k povodňovým ztrátám, než například lín obecný (*Tinca tinca*), amur bílý (*Ctenopharyngodon idella*), nebo štika obecná (*Esox lucius*) či okoun říční (*Perca fluviatilis*). Rovněž byly prokázány i rozdíly v „citlivosti“ mezi juvenilními a dospělými jedinci v rámci různých druhů. Ukázalo se, že vysoké srážky, či rozloha povodně v oblasti nemusí nutně znamenat vysoké ztráty rybí populace. Daleko negativnější efekt pak měly přívalové deště, či lokální záplavy s tzv. „train“ efektem. V průběhu zpracování dat byla pro porovnání ztrát v jednotlivých letech a u jednotlivých druhů využita analýza rozptylu. Konkrétněji šlo o jednofaktorovou a dvoufaktorovou analýzu rozptylu s pevnými efekty. Dále byl využit neparametrický Kruskalův-Wallisův test. K následnému vícenásobnému srovnání byl využit modifikovaný Tukeyův HSD test, spolu s post-hoc Kruskalovo-Wallisovo testem. Výsledky konfirmačních analýz byly následně vhodně doplněny i o vizualizaci prostřednictvím skupinových box-whiskers diagramů dostupných prostřednictvím softwarového vybavení. K numerickému vyhodnocení byl využit komerční software **Statistica 12 CZ** firmy StatSoft, Inc. ■

V letech 2000 až 2005 autor spolupracoval s významnou americkou generickou firmou IVAX, patřící v té době mezi 3 největší generické výrobce na světě, na klinickém hodnocení mikroemulzního preparátu s obsahem cyclosporinu (generické léčivo s obsahem imunosupresiva). Výsledky klinického

hodnocení jsou součástí registrační dokumentace, která posloužila k registraci léku v USA v roce 2005. V období 2006 až 2009 autor spolupracoval s největší rodinnou farmaceutickou společností v Latinské Americe, EMS Brazílie a firmou Transmedical se sídlem v Libanonu. Cílem spolupráce bylo statistické zpracování klinických studií. Výsledky byly součástí registrační dokumentace, na základě které byl lék zaregistrován v roce 2008 v EU. Část výsledků plynoucí z této spolupráce, které nejsou součástí registrační dokumentace, byla publikována během několika let v souboru prací, viz publikace [84], [85], [86] a v níže komentované práci [107].

Cílem práce [107] bylo porovnání farmakokinetiky a bioekvivalence nové generické formulace Cyclosporinu A (Equoral<sup>®</sup>) po převedení pacientů z původní originální formulace (Neoral<sup>®</sup>). Pomocí nerandomizované, fixní klinické studie bez opakování<sup>2</sup> byly naměřeny koncentrace CyA v krvi stabilizovaných pacientů. U 70 pacientů byly stanoveny různé charakteristiky pro farmakokinetiku obou preparátů pro posouzení bioekvivalence obou formulací. Některé zjištěné farmakokinetické charakteristiky byly vzhledem k zeshiknutí logaritmičtě transformovány. Následně byla testována bioekvivalence obou přípravků pomocí testů ekvivalence (tzv. two-one-sided *t*-test) a analýzy rozptylu (pro posouzení vlivu center ve kterých byla studie realizována). Rovněž byly zkonstruovány i 90 % nejkratší konfidenční intervaly pro podíl mediánů originátora a generika. Ukázalo se, že mezi formulacemi nebyl z hlediska bioekvivalence detekován signifikantní rozdíl. ■

V následujících pracích (komentář ukončený symbolem ★) byly využity především regresní metody, ať již lineární či nelineární regresní modely, viz

---

<sup>2</sup>tzv. „non-randomized, steady-state clinical study with fixed non-replicate study design“. Generická léčiva, léčiva identická s etickými léky, kterým vypršela patentová ochrana, musí splňovat kritéria stanovená národními registračními úřady. Jedná se o soubor požadavků na fyzikální, chemické a biologické vlastnosti, které v sumě zajišťují, že se daný přípravek shoduje se svým etickým originálem. Vedle požadavků na kvalitu, množství, příp. krystalickou strukturu účinné látky, stabilitu, rozpadavost, in vitro disoluční profil léčiva ve vybraných pufrch apod. je zásadním požadavkem pro registraci generického léčiva shodné chování v těle pacientů. K porovnání profilů etického a generického léčiva v lidském těle slouží bioekvivalenční studie. Jedná se o kontrolovanou studii, kdy se vybrané kohortě dobrovolníků, případně stabilizovaných pacientů, náhodně rozdělené do dvou skupin, jednorázově zkříženě podá stejná dávka referenčního (etického) a testovaného (generického) léčiva. Ve stanových časových intervalech se dobrovolníkům odebírají tělesné tekutiny (krev, sérum, moč), které slouží ke stanovení obsahu účinné látky v různých časových intervalech od podání léčiva. Vzhledem k tomu, že se jedná o zkříženou studii, získáváme od každého dobrovolníka informace o chování jak etického, tak generického léčiva.

[136]. Případně v nich byla využita metodologie lineárních smíšených modelů, viz publikace [87] nebo zobecněných lineárních modelů, viz diskutovaná práce [102], které umožňují modelovat například binární závisle proměnnou. Diskutována je zde i práce, kde je využita metodologie zobecněných smíšených lineárních modelů, viz [99]. Je třeba říci, že by bylo možné do této partie zařadit i výše diskutovanou publikaci [68], ve které je využita analýza přežívání, neboť ta je svým způsobem rovněž zvláštním případem „regresního“ modelu.

V publikaci [102] byl mimo jiné studován vliv věku zvířat, technologie chovu a odstavení na výskyt kryptosporidiových infekcí v chovech prasat (*Sus scrofa*). S cílem zjistit, zda výše uvedené faktory ovlivňují výskyt kryptosporidií a s ohledem na povahu modelované závislé proměnné, přítomnost/nepřítomnost infekce byl využit zobecněný lineární model s logistickou linkovací funkcí.

Vzhledem k získaným hodnotám logistických regresních koeficientů odhadnutého GLM modelu se ukázalo, že s rostoucím věkem u zvířat dochází k signifikantnímu poklesu výskytu druhu *Cryptosporidium suis*, zatímco věk neměl vliv na výši prevalence druhu *C. scrofarum* na hladině významnosti  $\alpha = 0,05$ . Dalším dílčím výsledkem bylo zjištění, že výskyty infekce způsobené *C. suis* byly pravděpodobnější u selat do pěti týdnů věku, ve srovnání s ostatními prasaty. Naopak se ukázalo, že selata starší šesti týdnů byla častěji infikována *C. scrofarum*. V následujících experimentech, viz [72], byla prokázána věková specifita *C. scrofarum*, tedy, že tento druh není infekční pro selata mladší pěti týdnů věku. Tím byly potvrzeny výše uvedené výsledky.

Další výsledky [102] neprokázaly vliv věku zvířat při odstavení na výskyt *C. suis*. Zvířata odstavená ve třech týdnech věku byla 2× častěji infikována druhem *C. scrofarum*. Použitá logistická regrese prokázala pozitivní efekt pozdějšího věku při odstavení na výskyt *C. scrofarum*. V neposlední řadě byl rovněž prokázán vliv technologie chovu na výskyt sledovaných patogenů. Zvířata chovaná na hluboké podestýlce byla přibližně 3× častěji infikována než ta, která byla ustájena na betonových nebo roštových podlahách.

Dnes by bylo zajímavé přistoupit k výše zmíněné analýze prostřednictvím bayesovského přístupu a pokusit se stanovit posteriorní distribuce jednotlivých logistických regresních koeficientů odhadnutého GLM modelu, například pomocí software WinBUGS, R, nebo STAN. ★

V publikaci [136] byly studovány kvalitativní změny uskladněného rybiho masa (*Cyprinus carpio* L.) v průběhu 84 dnů, při uskladnění v  $-20^{\circ}\text{C}$ . Pomocí instrumentální texturní analýzy byla sledována kvalita masa prostřednictvím tuhosti, soudržnosti a elasticity. Ukazatele byly měřeny v Newtonech či bezrozměrných jednotkách. K modelování změn byla využita klasická regresní analýza. K popisu změn se v případě elasticity ukázal jako vyhovující prostý lineární model.

U tuhosti a soudržnosti byl naopak identifikován značně nelineární průběh změn kvality masa. Vzhledem nelinearitě, resp. s ohledem na tvar korelačního pole byl pro obě veličiny využit model negativního exponenciálního poklesu, který dobře aproximoval tvar korelačního pole. Navržený model byl před vlastním numerickým odhadem jednotlivých regresních parametrů parametrizován tak, aby jeho jednotlivé parametry měly i vhodnou, „praktickou“ interpretaci.

Systematickou část modelu tak bylo možné vyjádřit následovně:  $\theta_0 + \theta_1 2^{-t/\theta_2}$ . K získání odhadů parametrů modelu byl využit iterační Gaussův-Newtonův algoritmus [6]. Numerický výpočet byl realizován prostřednictvím programovacího prostředí R při využití funkce `nls` [137]. ★

Práce [99] se zaměřila na srovnání výskytu patogenů, které jsou obsaženy v trusu hostitelů (*Pongo abelii* a *Pongo pygmaeus*) a jsou vylučováni nepravidelně v rozmezí 2 – 14 dnů, přičemž není vyrovnaný počet sledovaných vzorků u jednotlivých vyšetřovaných skupin.

Při analýze dat byl použit zobecněný lineární smíšený model (GLMM) za předpokladu závisle proměnné s binomickou distribucí. Tou byla přítomnost/nepřítomnost parazita ve vzorcích trusu různých skupin orangutanů. GLMM model byl využit vzhledem k nevyváženosti odběru jednotlivých vzorků v průběhu sběru dat. Samotný model využíval jako pevné vysvětlující proměnné „ostrov“ (Kalimantan, Sumatra) a „úroveň kontaktu s lidmi“, která byla sledována na třech úrovních dle stupně kontaktu s lidmi. Náhodné efekty (lokalita a individuum) byly vnořeny do odpovídajících si fixních efektů. Potřebné výpočty byly realizovány opět pomocí programovacího prostředí R, konkrétněji pak pomocí knihovny `glmm`.

Výsledky studie ukázaly, že volně žijící orangutani jsou signifikantně méně často parazitováni kryptosporidii než jedinci, kteří jsou chováni v zajetí nebo jsou plně habituováni. Zároveň byl prokázán vliv prostředí na výskyt sledovaných kryptosporidií. Naopak nebyl zjištěn vliv habituace a prostředí na výskyt některých mikrosporidiových infekcí (*Encephalitozoon spp.* a *Enterocytozoon bieneusi*). ★

Práce [87] se zaměřuje na studium vlivu příkrmování, prostřednictvím upravených obilovin (tritikále, ječmen, žito a lisovaná pšenice v kombinaci s řepkou), na přírůstky kapra (*Cyprinus carpio* L.) v poloprovozních podmínkách.

V průběhu experimentů byly využity dvě skupiny, u nichž byla realizována pouze přirozená potrava - zooplankton, které následně sloužily jako kontrolní skupiny. Před vlastním modelováním byla ověřena na začátku poloprovozních pokusů homoskedasticita hmotnosti v jednotlivých skupinách (rybí obsádky) a to prostřednictvím klasického Bartlettova testu. V případě plánovaných párových srovnání byl použit dvouvýběrový  $F$ -test. Pro konstrukci konfidenčních intervalů směrodatných odchylek byl využit klasický asymptotický přístup.

Vzhledem k zaměření pokusu na čtvrté růstové období, byly pro modelování závislosti za adekvátní zvoleny lineární funkce a to i pro jejich relativně snadnou interpretaci a skutečnost, že v průběhu experimentu vykazovala růstová trajektorie zhruba konstantní průběh (lineární trend). Konkrétněji byla pro modelování závislosti využita metodologie lineárních smíšených modelů.

Celkem bylo navrženo pět lineárních smíšených modelů, které bylo možno ve shodě s literaturou [109] zapsat v maticové formě jako  $y_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$ , kde vektor  $\mathbf{b}_i \sim \mathbf{N}(0, \boldsymbol{\Psi})$  a vektor  $\boldsymbol{\epsilon}_i$  představoval vnitroskupinovou náhodnou chybovou složku, která sleduje  $\mathbf{N}(0, \sigma^2; \boldsymbol{\Lambda}_i)$ . Uvažované lineární modely bylo možno popsat následovně: 1) lineární smíšený model s náhodnou absolutní složkou (tzv. random intercept model), 2) lineární smíšený model s náhodným lineárním členem (tzv. random slope model), 3) lineární smíšený model s náhodnou absolutní i lineární složkou. Zbylé dva smíšené modely (modely 4 a 5) do lineárního modelu přidávaly navíc složku umožňující zachytit i různou variabilitu chybové složky  $\epsilon_{ij}$ , přičemž poslední model umožňoval zohlednit i různé druhy krmení. Tento model se ukázal jako nejvhodnější pro popis růstu ve sledovaném období.

K samotnému odhadu regresních parametrů byla použita metoda maximální věrohodnosti, blíže viz [109]. Numerické výpočty byly realizovány prostřednictvím programovacího prostředí R a knihovny nlme autorů Pinheira, Batese a kol. [110]. Vhodnost jednotlivých modelů byla posuzována prostřednictvím AIC a BIC kritérií.

Hlavní výsledky provedených poloprovozních pokusů ukazují na skutečnost, že růst, potažmo přírůstky, mohou být do jisté míry ovlivněny i úpravou

krmiv, neboť lisované žito, nebo ječmen, dosahovalo lepších výsledků a tedy i snížení provozních nákladů v porovnání s ostatním krmivem. Rovněž lze říci, že lisovaná krmiva vykazovala spíše nižší variabilitu, resp. heteroskedasticitu růstových trajektorií, a to ve srovnání s ostatními, tlakem neupravenými krmivy.

Z dnešního pohledu by mohlo být zajímavé uplatnění kvantilové regrese, při hodnocení růstu u různých skupin. Tímto způsobem by mohlo být získáno mnohem více informace o průběhu růstu v různých „frakcích“ rybí obsádky, v rámci jednotlivých experimentálních skupin. Rovněž by mohlo být zajímavé i uplatnění funkcionální analýzy dat, která by mohla, v případě aplikace splinového vyhlazování s různým řádem a adekvátně odhadnutým penalizačním parametrem, poskytnout i detailnější vhled do dynamiky růstu kapra obecného v různých poloprovozních podmínkách, což by mohlo mít i další dopad do praxe. ★

V průběhu několika posledních let autor úzce spolupracoval s firmou Gemini PharmChem České Budějovice a.s., která je dceřinou společností švýcarské společnosti Synbias AG. Zde se autor podílel na vývoji finálních léčivých forem a na optimalizaci biotechnologických procesů. Jedním z několika dosavadních výsledků byla i nová formulace kladribinu. Ta se stala předmětem podaného evropského patentu *EP18196664.9*. V průběhu vývoje nové nekolizní formulace kladribinu se uplatnila především metodologie plánování experimentů spolu s analýzou responsních povrchů a následným využitím optimalizačních metod založených na přírodou inspirovaných heuristikách. (komentář k tomuto výsledku je ukončen symbolem ▲)

Kladribin<sup>3</sup>, chemicky 2-chlor-2'-deoxyadenosin (2CdA), je purinový analog, původně vyvinutý a používaný při léčbě zejména hematologických onemocnění. První registrace parenterální formulace byla udělena již v roce 1993. Zvýšený zájem o kladribin vyvolaly informace o jeho možném využití při léčbě roztroušené sklerózy (RS).

Pro indikaci v léčbě RS bylo nutné vyvinout orální formu léčiva, kterou jako

---

<sup>3</sup>Kladribin je definován jako léčivo s rychlou rozpustností a nízkou permeabilitou (třída III, podle BCS, Biophaceutical Classification System, klasifikace), což představuje velkou výzvu pro vývoj orální formulace. Navíc je kladribin velmi citlivý k nízkému pH. V pH 1.5 se rychle rozkládá a po hodinové expozici se v roztoku nachází asi 5-10% původní molekuly. Podle literatury neexistuje specifický transportní systém pro kladribin a látka se vstřebává pasivní difuzí.

první patentovala firma IVAX v roce 2003. Formulace IVAX byla založena na využití cyklodextrinu jako cukerného nosiče, s deklarovanou schopností tvořit komplex s kladribinem a chránit jeho molekulu před rozkladem v nízkém pH. Formulace podle původního patentu je velmi komplikovaná jak na přípravu, tak na čas. Je založena na použití vysokých koncentrací cyklodextrinů ( $\beta$ -cyklodextrinu a jeho derivátů), dlouhém čase míchání a finální lyofilizaci premixu.

Cílem našeho týmu bylo vyvinout nekolizní formulaci, která by odstranila komplikovanou technologii přípravy a zároveň zjednodušila a zlevnila proces přípravy léčiva. Vzhledem k tomu, že kladribin není transportován do buněk aktivně, rozhodli jsme se pro formulaci, která umožní rychlé uvolnění účinné látky v žaludku a následné zvýšení koncentrace léčiva. V důsledku zvýšení koncentrace se zvýší i absorpce léčiva. Podmínkou bylo, aby absorpce léčiva v nízkém pH proběhla rychleji, než jeho rozklad a aby bylo dosaženo dostatečně vysokých hodnot účinné látky v organismu.

Vzhledem k tomu, že rozpustnost látek ovlivňuje jejich krystalická struktura, rozhodli jsme se využít amorfní formu léčiva, která je sice nejméně stabilní, ale vykazuje nejvyšší rozpustnost.

Dalším důležitým krokem byl výběr vhodného nosiče léčiva. Zjistili jsme, že nerozpustné nosiče na bázi celulózy a jejich derivátů, běžně využívané pro přípravu premixů, nejsou vhodné pro dosažení rychlé rozpustnosti formulace. Nakonec byl vybrán zcela nový, komerčně dostupný, fruktózový oligosacharid, linecap (Roquette, Francie), který poskytoval jak stabilní, tak rychlorozpustný premix vhodný pro přípravu orální lékové formy s obsahem kladribinu.

V průběhu různých etap vývojových prací se jako výhodné ukázalo využití výše popsané metodologie DOE a RSM, spolu s následným využitím metod pro optimalizaci responsních povrchů, při existenci omezení zapříčiněných fyzikálně-chemickými vlastnostmi směsi. Pro dosažení rychlé rozpustnosti navrhované formulace byly využity „alfabeticky“ optimální plány a to s ohledem na výše zmiňovaná omezení. Numerické výpočty se realizovaly v prostředí R za pomoci knihoven MASS [137], mixexp [78], rsm [118], base [111], desirability [74], GA [125], [126] a DEoptim [96]. Při sestavování skriptu pro optimalizaci formulace, prostřednictvím knihoven GA a DEoptim, byly účelové funkce autorem naprogramovány tak, aby byla preferována řešení, vyhovující omezujícím podmínkám, a naopak penalizována řešení, která podmínky porušovala. ▲

Publikace [95] navrhuje nový statistický test využitelný při analýze funkcionálních dat. Práce byla v únoru 2019 zaslána k oponentnímu řízení do časopisu „Methodology & Computing in Applied Probability“ (komentář k této podané práci je ukončen symbolem ◀). Oponentní řízení zatím stále trvá.

Námi navržená neparametrická procedura umožňuje testovat hypotézu  $H_0 : \beta_{jk} = 0$ , tj. významnost faktorů prostřednictvím tzv. „rank envelop testu“ v případě funkcionálního obecného lineárního modelu. Ve své podstatě jede o rozšíření metody popsané Mrkvičkou a kolektivem [94] na obecnější případ.

Výhodou navrženého testu je především jeho grafická interpretace. Test umožňuje nejenom identifikovat zda je daný faktor významný, ale navíc umožňuje identifikovat i oblast funkcionálních dat, která způsobuje zamítnutí nulové hypotézy. V případě funkcionální vícefaktorové analýzy rozptylu hlavních efektů nebo funkcionální analýzy kovariance hlavních efektů je test schopen zjistit, které skupiny se liší a v jaké části – oblasti – funkce k tomu dochází. Procedura je vzhledem k použití Freedmanova–Laneova permutačního schématu, viz [42], mírně liberální.

Na druhou stranu lze na základě provedených počítačových simulací říci, že námi navržený test je mnohem silnější, než tzv. „random projection method“ či „ $F$ -max“ metoda. Tato výhoda se navíc výrazněji projevuje u procesů, které nejsou i.i.d. ◀



## Závěr a přínos habilitační práce pro obor

Habilitační práce podává přehled různých statistických metod aplikovatelných při plánování, analýze a interpretaci experimentálních dat plynoucích z biologických experimentů s odkazy na vybrané a autorem ve spoluautorství publikované práce. Na příkladech publikovaných výsledků pak byla demonstrována jejich praktická využitelnost.

Je zcela nezpochybnitelné, že správné plánování prováděných experimentů a správné zpracování a interpretace zásadním způsobem ovlivňuje finální výstup experimentálních studií. Vzhledem k současnému stavu poznání v různých vědeckých disciplínách je nutná spolupráce odborníků z různých vědních oborů.

Dnes není prakticky možné, aby jednotlivec obsáhl tak odlišné vědní obory, jakými jsou biologie, zootechnika, mikrobiologie a biotechnologie na jedné straně a kvantitativní metody zahrnující aplikovanou statistiku, optimalizaci, numerické metody či programování v různých jazycích na straně druhé. Proto je žádoucí, aby součástí vědeckého týmu byli jak biologové, tak i odborníci na plánování experimentů, analýzu a interpretaci zpracovávaných experimentálních dat.

Většina, zejména *in vivo* experimentů představuje unikátní soubory často získané s velkou obětavostí a zájmem, mnohdy těžko opakovatelné soubory (počty zvířat, délka trvání experimentů, etické požadavky) a bylo by chybou, aby například nevhodným naplánováním experimentů či jejich nevhodným vyhodnocením došlo ke „ztrátám“ či zkreslení získaných experimentálních výsledků nebo takřkajíc nezískání „možného“.

Domnívám se, že použití vhodné statistické metodologie uvedené v předkládané habilitační práci může přispět ke zkvalitnění experimentální práce a rovněž umožnit získání komplexnějších informací o studované problematice. Tou může být experiment zaměřený na přírůstky u hospodářských zvířat, kvalitu masa nebo genomiku či optimalizaci biotechnologického procesu.

# Implementace dosažených poznatků ve vysokoškolské výuce

Autor předložené práce byl školitelem šesti bakalářských prací a pěti diplomových prací, které byly zaměřeny na využití vícerozměrných statistických metod, klasifikačních a data miningových metod, či případně na metodologii plánování experimentů. Autor rovněž působil jako konzultant několika diplomových a bakalářských prací na různých součástech Jihočeské univerzity.

Během své akademické praxe autor působil na třech fakultách JU v Českých Budějovicích: ZF (roky: 2002 - 2006, 2019 - dosud), EF (roky: 2007 - 2019), ZSF (roky: 2003 - 2016) při JU v Českých Budějovicích. Zde je či byl garantem několika předmětů. Za všechny lze jmenovat především: Teorii pravděpodobnosti a statistiky, Vyrovnávací seminář z pravděpodobnosti, Statistika I a II., Biostatistika.

V současné době autor pracuje na Katedře genetiky a speciální produkce rostlinné ZF JU v Českých Budějovicích, spolupracuje s pracovištěm Laboratorní analytické biochemie a metabolomiky Entomologického ústavu při Biologickém centru AVČR v. v. i. Částečně stále spolupracuje s Katedrou aplikované matematiky a informatiky EF JU v Českých Budějovicích, kde byl zaměstnán od jejího vzniku, až do roku 2019. V současnosti autor připravuje na Zemědělské fakultě v Českých Budějovicích předměty Biostatistika I, Biostatistika II a Pravděpodobnost a statistika, které se na ZF JU v Českých Budějovicích stanou součástí nově akreditovaných studijních programů.

Autor je rovněž jedním ze spoluautorů knihovny GET pro programovací prostředí R umožňující provádět mimo jiné, např. dvoufaktorovou analýzu roz-

ptylu pro funkcionální data [95]<sup>1</sup> prostřednictvím konceptu globálních obálek, viz [97]. Knihovna `GET` je od července 2019 dostupná na Comprehensive R Archive Network programovacího prostředí R a může být využívána širokou odbornou veřejností. Autor současně působí již více než patnáct let v komerční sféře, kde se specializuje na konzultační činnost související s aplikovanou statistikou, vyhodnocováním bioekvivalenčních studií, plánováním experimentů a optimalizací biotechnologických procesů.

---

<sup>1</sup>Práce je v současnosti v oponentním řízení, viz výše.

## Publikační činnost autora

Níže je uveden seznam vybraných vědeckých prací s IF na nichž se autor habilitační práce podílel. U jednotlivých prací je uváděna poslední dostupná hodnota IF dle WoS.

- [54] Holubová, N., Sak, B., Hlásková, L., Kvetoňová, D., Hanzal, V., Rajský, D., Rost, M., McEvoy, J. Kváč, M. Host specificity and age-dependent resistance to *Cryptosporidium avium* infection in chickens, ducks and pheasants. *Experimental Parasitology*, 2018, 191, 8, 62-65. IF 1,821
- [68] Kváč, M., Kodádková, A., Sak, B., Květoňová, D., Jalovecká, M., Rost, M., Salát, J. Activated CD8+ T cells contribute to clearance of gastric *Cryptosporidium muris* infections. *Parasite Immunology*, 2011, 33, 4, 210-216. IF 2,836
- [69] Kváč, M., Hromadová, N., Květoňová, D., Rost, M., Sak, B. Molecular characterization of *Cryptosporidium* spp. in pre-weaned dairy calves in the Czech Republic: Absence of *C. ryanae* and management-associated distribution of *C. andersoni*, *C. bovis* and *C. parvum* subtypes. *Veterinary Parasitology*, 2011, 177, 3/4, 378-382. IF 2,422
- [70] Kváč, M., McEvoy, J., Loudová, M., Stenger, B., Sak, B., Květoňová, D., Ditrich, O., Rašková, V., Moriarty, E., Rost, M., Macholan, M., Piálek, J. Coevolution of *Cryptosporidium tyzzeri* and the house mouse (*Mus musculus*). *International Journal for Parasitology*, 2013, 43, 10, 805-817. IF 3,078
- [71] Kváč, M., Ondráčková, Z., Květoňová, D., McEvoy, J., Vítovec, J., Rost, M., Sak, B. The Lesser Egyptian Gerbil (*Gerbillus gerbillus*) is a suitable host for the long-term propagation of *Cryptosporidium andersoni*. *Experimental Parasitology*, 2013, 134, 4, 438-442. IF 1,821

- [72] Kváč, M., Němejc, K., Kestránová, M., Květoňová, D., Wágnerová, P., Kotková, M., Rost, M., Samková, E., McEvoy, J., Sak, B. Age related susceptibility of pigs to *Cryptosporidium scrofarum* infection. *Veterinary Parasitology*, 2014, 202, 3-4, 330-334. IF 2,422
- [76] Laatamna, A. E., Wágnerová, P., Sak, B., Květoňová, D., Aissi, M., Rost, M., Kváč, M. Equine cryptosporidial infection associated with *Cryptosporidium* hedgehog genotype in Algeria, *Veterinary Parasitology*, 2013, 197, 1-2, 350-353. IF 2,422
- [77] Laatamna, A. E., Wágnerová, P., Sak, B., Květoňová, D., Xiao, L., Rost, M., McEvoy, J., Saadi, Ar., Aissi, M., Kváč, M. Microsporidia and *Cryptosporidium* in horses and donkeys in Algeria: Detection of a novel *Cryptosporidium hominis* subtype family (Ik) in a horse. *Veterinary Parasitology*, 2015, 208, 3-4, 135-142. IF 2,422
- [84] Masri, M.A., Haberal, M. Rizvi, A., Stephan, A., Bilgin, N., Naqvi, A., Barbari, A., Kamel, G., N. Zafar, Emirog Lu, R., Colak, T., Manzoor, K., Matha, V., Kamarád, V., Rost, M., Rizk, S., Hazime, A., Perlík, F. Switchability of Neoral and Equoral According to Food and Drug Administration Rules and Regulations. *Transplantation Proceedings*. New York : Elsevier Science Inc., 2005, 37, 2988-2993. ISSN 0041-1345. IF 0,511
- [85] Masri, M., Rizk, S., Barbari, A., Stephan, A., Kamel, G., Rost, M. An assay for the determination of sirolimus levels in the lymphocyte of transplant patiens. *Transplantation Proceedings*, 2007, 39, 1204-1206. IF 0,806
- [86] Masri, M. Rizk, S., Attia, M., Barbouch, H., Rost, M. Bioavailability of a New Generic Formulation of Mycophenolate Mofetil MMF 500 Versus CellCept in Healthy Adult Volunteers. *Transplantation Proceedings*. New York : Elsevier Science Inc., 2007, 39 (4), 1233-1236. ISSN 0041-1345. IF 0,511
- [87] Másílko, J., Hartvich, P., Rost, M., Urbánek, M., Hlaváč, D., Dvořák, P. Potential for Improvement of Common Carp Production Efficiency by Mechanical Processing of Cereal Diet. *Turkish Journal of Fisheries and Aquatic Sciences*, 2014, 14, 1, 145-153. IF 0,482
- [99] Mynářová, A., Foitová, I., Kváč, M., Květonová, D., Rost, M., Morrogh-Bernard, H., Nurcahyo, W., Nguyen, C., Supriyadi, S., Sak, B. Pre-

valence of *Cryptosporidium* spp., *Enterocytozoon bienersi*, *Encephalitozoon* spp. and *Giardia intestinalis* in Wild, Semi-Wild and Captive Orangutans (*Pongo abelii* and *Pongo pygmaeus*) on Sumatra and Borneo, Indonesia. Plos ONE, 2016, roč. 11, č.3. IF 2,766

- [102] Němejc, K, Sak, B., Květoňová, D., Kernerová, N., Rost, M., Cama, VA., Kváč, M. Occurrence of *Cryptosporidium suis* and *Cryptosporidium scrofarum* on commercial swine farms in the Czech Republic and its associations with age and husbandry practices Parasitology Research, 2013, 112, 3, 1143-1154. IF 2,558
- [106] Ondráčková, Z., Kváč, M., Sak, B., Květoňová, D., Rost, M. Prevalence and molecular characterization of *Cryptosporidium* spp. in dairy cattle in South Bohemia, the Czech Republic. Veterinary Parasitology. Elsevier B.V., 2009, roč. 165, 1-5, 141-144. IF 2,422
- [107] Perlík, F., Masri, M.A., Rost, M., Kamarád, V. Pharmacokinetic conversion study of a new cyclosporine formulation in stable adult renal transplant recipients. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2005, 149, 2, 309-13. IF 1,141
- [119] Rutkayová, J., Vácha, F., Maršálek, M., Beneš, K., Civišová, H., Horká, P., Petrášková, E., Rost, M., Šulista, M. Fish stock losses due to extreme floods - findings from pond-based aquaculture in the Czech Republic. Journal of Flood Risk Management, 2018, 11, 3, 351-359. IF 2,483
- [120] Rutkayová, J., Voříšková, J., Beneš, K., Hasoňová, L., Vácha, F., Křížek, M., Rost, M., Jirotková, D., Hanusová, L. The "Freshness" Objective Assessment of Carp Meat - Is It Appropriate to Use Freshness Meter? Chemické listy, 2018, 112, 5, 333-339 IF 0,26
- [122] Sak, B., Kváč, M., Květoňová, D., Kučerová, Z., Rost, M., Secor, E. Seropositivity for *Enterocytozoon bienersi*, Czech Republic. Emerging Infectious Diseases, 2010, 16, 2, 335-337. IF 7,422
- [123] Sak, B., Brady, D., Pelikánová, M., Květoňová, D., Rost, M., Kostka, M., Tolarová, V., Hůzová, Z., Kváč, M. Unapparent Microsporidial Infection among Immunocompetent Humans in the Czech Republic.. Journal of Clinical Microbiology, 2011, 49, 3, 1064-1070. IF 4,054
- [124] Sak, B., Jandová, A., Doležal, K., Kváč, M., Květoňová, D., Hlásková, L., Rost, M., Olšanský, M., Nurcahyo, W., Foitová, I. Effects of selected

- Indonesian plant extracts on *E. cuniculi* infection in vivo. *Experimental Parasitology*, 2017, 181, 10, 94-101. IF 1,821
- [129] Slavíková, L., Malý, V., Rost, M., Petružela, L., Vojáček, O. Impacts of Climate Variables on Residential Water Consumption in the Czech Republic. *Water Resources Management*, 2013, 27, 2, 365-379. IF 2,6
- [135] Urbánek, M., Hartvich, P., Vácha, F., Rost, M. Investigation of fat content in market common carp (*Cyprinus carpio*) flesh during the growing season. *Aquaculture nutrition*, 2010, 16, 5, 511-519. IF 0.83
- [136] Vácha, F., Cepák, M., Urbánek, M., Vejsada, P., Hartvich, P., Rost, M. Impact Of Long-Term Storage On The Instrumental Textural Properties Of Frozen Common Carp (*Cyprinus Carpio*, L.) Flesh. *International Journal of Food Properties*, 2013, 16, 2, 241-250. IF 0,915
- [140] Wagnerová, P., Sak, B., McEvoy, J., Rost, M., Matysiak, A., Ježková, J., Kváč, M. Genetic diversity of *Cryptosporidium* spp. including novel identification of the *Cryptosporidium muris* and *Cryptosporidium tyzzeri* in horses in the Czech Republic and Poland. *Parasitology Research*, 2015, 114, 4, 1619-1624. IF 2,558
- [141] Wágnerová, P., Sak, B., McEvoy, J., Rost, M., Sherwood, D., Holcomb, K., Kváč, M. *Cryptosporidium parvum* and *Enterocytozoon bieneusi* in American Mustangs and Chincoteague ponies. *Experimental Parasitology*, 2016, 162, 1, 24-27. IF 1,821



## Literatura

- [1] Aggarwal, Ch. C. A Reddy, Ch. K. Data clustering: algorithms and applications, 2014, Boca Raton: Chapman and Hall/CRC.
- [2] Agresti A. Categorical Data Analysis, Wiley Series in Probability and Statistics, 3rd edition, 2003, Wiley Interscience John Wiley & Sons, New York.
- [3] Anderson, M. J., Whitcomb, P. J., Bezener, M. A. Formulation simplified: finding the sweet spot through design and analysis of experiments with mixtures, 2018, New York: Taylor & Francis.
- [4] Anděl, J. Základy matematické statistiky, 3 vyd., 2011, Praha: Matfyzpress.
- [5] Basilevsky, A. Statistical factor analysis and related methods: theory and applications, 1994, New York: Wiley.
- [6] Bates, D. M., Watts, D. G. Nonlinear Regression Analysis and Its Applications, 1988, Wiley.
- [7] Bartel, J., Krumsiek, J., Theis, F. J. Statistical methods for the analysis of high-throughput metabolomics data, Computational and structural biotechnology journal, 2013, 4.
- [8] Bi, J., Kuesten, C. Using Functional Data Analysis (FDA) Methodology and the R Package „FDA“ for Sensory Time-Intensity Evaluation, Journal of Sensory Studies, 2013, 28, 6, 474–482.
- [9] Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, Journal of Chemometrics, 2006, 20, 8-10, 341–351.

- [10] Bilder, C. R., Loughin, T. M. Analysis of categorical data with R, 2015, Boca Raton: CRC Press, Taylor & Francis Group.
- [11] Boleda, M. D., Briones, P., Farrés, J., Tyfield, L., Pi, R. Experimental Design: A Useful Tool for PCR Optimization, *BioTechniques*, 1996, 21,1, 134 - 140.
- [12] Borg, I., Groenen, P. J. F. Modern multidimensional scaling: theory and applications, 2nd ed, 2005, New York: Springer.
- [13] Box, G. E. P., Hunter, J. S. The  $2^{k-p}$  fractional factorial designs, *Technometrics*. 1961, 3, 311–351, 449–458.
- [14] Box, G. E. P., Draper, N. R. Empirical model-building and response surfaces, 1987, New York: Wiley.
- [15] Box, G. E. P., Hunter, J. S., Hunter, W. G. Statistics for experimenters: design, innovation, and discovery. 2nd ed., 2005, Hoboken, N.J.: Wiley-Interscience.
- [16] Box, G. E. P., Draper, N. R. Response surfaces, mixtures, and ridge analyses. 2nd ed., 2007, Hoboken, N. J.: John Wiley.
- [17] L. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. Classification and Regression Trees, 1984, Chapman & Hall, New York.
- [18] Breiman, L. Bagging predictors. *Machine Learning*, 1996, 26, 2, 123–140.
- [19] Breiman, L. Random Forest, *Machine Learning*, 2001, 45, 1, 5-32.
- [20] Byrd, R. H., Lu, P., Nocedal, J., Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 1995, 16, 5, 1190–1208.
- [21] Campisi, B., Chicco, D., Vojnovic, D., Phan-Tan-Luu, R. Experimental design for a pharmaceutical formulation: optimisation and robustness. *Journal of Pharmaceutical and Biomedical Analysis*, 1998, 18, 1-2, 57–65.
- [22] Cobb, B. D., Clarkson, J. M. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods, *Nucleic Acids Res*, 1994, 22, 18, 3801-3805.

- [23] Cornell, J. A. Experiments with mixtures: designs, models, and the analysis of mixture data. 3rd ed., 2002, New York: Wiley.
- [24] Cornell, J. A. A primer on experiments with mixtures, 2011, Hoboken, N. J.: Wiley, Wiley series in probability and statistics.
- [25] N. Cristianini, N., Shawe-Taylor, J. An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods, 2000, Cambridge University Press, New York.
- [26] Davison, A. C., Hinkley, D. V. Bootstrap Methods and their Application, Cambridge Series in Statistical and Probabilistic Mathematics, 1997, Cambridge University Press.
- [27] De Aguiar, P. F., Bourguignon, B., Khots, M. S., Massart, D. L., Phan-Thau-Luu, R. D-optimal designs. Chemometrics and Intelligent Laboratory Systems, 1995, 30, 2, 199 - 210.
- [28] Draper, N. R., Smith, H. Applied regression analysis, 3rd ed., 1998, New York: Wiley.
- [29] De Baun, R. Response Surface Designs for Three Factors at Three Levels. Technometrics, 1959, 1, 1, 1-8.
- [30] Derringer, G., Suich, R. Simultaneous Optimization of Several Response Variables. Journal of Quality Technology, 1980, 12, 4, 214–219.
- [31] Duda, R. O., Hart, P. E., Stork, D. H. Pattern Classification, 2nd ed., 2000, Wiley Interscience.
- [32] Duong, T. ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. Journal of Statistical Software, 2007, 21, 7, 1-16.
- [33] Efron, B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation, Journal of the American Statistical Association, 1983, 78, 382, 316-331.
- [34] Efron, B., Tibshirani, R. J. An introduction to the bootstrap. Monographs on Statistics and Applied Probability, 1993, Chapman and Hall, New York.
- [35] Ekpenyong M. G., Antai S. P., Asitok A. D., Ekpo B. O. Plackett-Burman Design and Response Surface Optimization of Medium Trace

Nutrients for Glycolipopeptide Biosurfactant Production. *Iran Biomed J.*, 2017, 21, 4, 249–260.

- [36] El-Naggar, N. EA., El-Shweihy, N. M., El-Ewasy, S. M. Identification and statistical optimization of fermentation conditions for a newly isolated extracellular cholesterol oxidase-producing *Streptomyces cavourensis* strain NEAE-42, *BMC Microbiology*, 2016, 16, 217.
- [37] Ferraty, F., Vieu, P. P. *Nonparametric functional data analysis: theory and practice*, 2006, New York: Springer.
- [38] Fisher, R. A. *Statistical methods for research workers*. Genesis, Publishing Pvt Ltd, 1925.
- [39] Fisher, R. A. *The design of experiments*, 1935, Oliver and Boyd, Edinburgh.
- [40] Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, 7, 2, 179–188.
- [41] Fraley, C., Raftery, A. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, 2002, 97, 458, 611-631.
- [42] Freedman, D., Lane, D. A nonstochastic interpretation of reported significance levels, *Journal of Business & Economic Statistics*, 1983, 1, 4, 292-298
- [43] Freund, Y., Schapire, R. Experiments with a new boosting algorithm, *Machine Learning: Proceedings of the Thirteenth International Conference*, 1996, 148–156.
- [44] Friedman, J. H. Regularized Discriminant Analysis, *Journal of the American Statistical Association*, 1989, 84, 405, 165-175.
- [45] Georgiou, S. D. Supersaturated designs: A review of their construction and analysis. *Journal of Statistical Planning and Inference*, 2014, 144, 92–109.
- [46] Goos, P., Jones, B. *Optimal design of experiments: a case study approach*. Hoboken, N. J.: Wiley, 2011.
- [47] Greenacre, M. J., Blasius, J. *Multiple correspondence analysis and related methods*, 2006, Boca Raton: Chapman & Hall/CRC.

- [48] Greenacre, M. J. Correspondence analysis in practice, 2nd ed., 2007, Boca Raton: Chapman & Hall/CRC.
- [49] Haaland, P. D. Experimental design in biotechnology, 1989, New York: Marcel Dekker.
- [50] Harrington, D. P., Fleming, T. R. A class of rank test procedures for censored survival data. *Biometrika*, 1982, 69, 3, 553–566.
- [51] Hastie, T., Tibshirani R., Friedman J. H. The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations, 2001, New York: Springer-Verlag.
- [52] S original by Hastie, T., Tibshirani, R. Original R port by Leisch, F., Hornik, K., Ripley, B. D. mda: Mixture and Flexible Discriminant Analysis, 2017, R package version 0.4-10, <https://CRAN.R-project.org/package=mda>
- [53] Hilbe, J. M. Logistic Regression Models, 2009, Chapman & Hall/CRC Press.
- [54] Holubová, N., Sak, B., Hlásková, L., Kvetoňová, D., Hanzal, V., Rajský, D., Rost, M., McEvoy, J., Kváč, M. Host specificity and age-dependent resistance to *Cryptosporidium avium* infection in chickens, ducks and pheasants. *Experimental Parasitology*, 2018, 191, 8, 62 - 65.
- [55] Holland, J. H. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, 2nd ed., 1992, Cambridge, Mass.: MIT Press.
- [56] Quinlan, J. R. C 4.5: Programs for machine learning, 1993, CA: Morgan Kaufman, San Francisco.
- [57] James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning: with applications in R, 2013, New York: Springer.
- [58] Jiawei, H., Kamber, M., Pei, J. Data mining: concepts and techniques. 3rd ed., 2012, Burlington, MA: Elsevier.
- [59] Jolliffe, I. T. Principal component analysis, 2nd ed., 2002, New York: Springer.
- [60] Källén, A. Computational pharmacokinetics, 2008, Boca Raton: Chapman & Hall/CRC.

- [61] Kaufman, L., Rousseeuw, P. J. Finding groups in data: an introduction to cluster analysis, 1990, New York: Wiley.
- [62] Keogh E., Mueen A. Curse of Dimensionality. In: Sammut C., Webb, G. I. (eds) Encyclopedia of Machine Learning, 2011, Springer, Boston, MA.
- [63] Key, M. A tutorial in displaying mass spectrometry-based proteomic data using heatmaps. BMC Bioinformatics, 2012, 13, (Suppl 16), S-10.
- [64] Koenker, R. W., Bassett, G. W. Regression quantiles, *Econometrica*, 1978, 46, 33–50.
- [65] Kruskal, W. H. and Wallis, A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 1952, 47, 583–621.
- [66] Khuri, A. I., Cornell, J. A. Response surfaces: designs and analyses. 2nd ed., rev. and expanded, 1996, New York: Marcel Dekker.
- [67] Khuri, A. I. Response surface methodology and related topics, 2006, Hacensack, N.J.: World Scientific.
- [68] Kváč, M., Kodádková, A., Sak, B., Květoňová, D., Jalovecká, M., Rost, M., Salát, J. Activated CD8+T cells contribute to clearance of gastric *Cryptosporidium muris* infections. *Parasite Immunology*, 2011, 33, 4, 210-216.
- [69] Kváč, M., Hromadová, N., Květoňová, D., Rost, M., Sak, B. Molecular characterization of *Cryptosporidium* spp. in pre-weaned dairy calves in the Czech Republic: Absence of *C. ryanae* and management-associated distribution of *C. andersoni*, *C. bovis* and *C. parvum* subtypes. *Veterinary Parasitology*, 2011, 177, 3-4, 378-382.
- [70] Kváč, M., McEvoy, J., Loudová, M., Stenger, B., Sak, B., Květoňová, D., Ditrich, O., Rašková, V., Moriarty, E., Rost, M., Macholan, M., Piálek, J. Coevolution of *Cryptosporidium tyzzeri* and the house mouse (*Mus musculus*). *International Journal for Parasitology*, 2013, 43, 10, 805-817.
- [71] Kváč, M., Ondráčková, Z., Květoňová, D., McEvoy, J., Vítovec, J., Rost, M., Sak, B. The Lesser Egyptian Gerbil (*Gerbillus gerbillus*) is a suitable host for the long-term propagation of *Cryptosporidium andersoni*. *Experimental Parasitology*, 2013, 134, 4, 438-442.

- [72] Kváč, M., Němejc, K., Kestřánová, M., Květoňová, D., Wágnerová, P., Kotková, M., Rost, M., Samková, E., McEvoy, J., Sak, B. Age related susceptibility of pigs to *Cryptosporidium scrofarum* infection. *Veterinary Parasitology*, 2014, 202, 3-4, 330-334.
- [73] Kokoszka, P., Reimherr, M. *Introduction to functional data analysis*, 2017, Boca Raton: CRC Press, Taylor & Francis Group.
- [74] Kuhn, M. *desirability: Function Optimization and Ranking via Desirability Functions*. R package version 2. 1., 2016, <https://CRAN.R-project.org/package=desirability>.
- [75] Kumar, N., Hoque, M. A., Sugimoto, M. Robust volcano plot: identification of differential metabolites in the presence of outliers. *BMC Bioinformatics*, 2018, 19 ,1.
- [76] Laalamna, AE., Wágnerová, P., Sak, B., Květoňová, D., Aissi, M., Rost, M., Kváč, M. Equine cryptosporidial infection associated with *Cryptosporidium hedgehog* genotype in Algeria, *Veterinary Parasitology*, 2013, 197, 1-2, 350-353.
- [77] Laalamna, AE., Wágnerová, P., Sak, B., Květoňová, D., Xiao, L., Rost, M., McEvoy, J., Saadi, AR., Aissi, M., Kváč, M. Microsporidia and *Cryptosporidium* in horses and donkeys in Algeria: Detection of a novel *Cryptosporidium hominis* subtype family (Ik) in a horse. *Veterinary Parasitology*, 2015, 208, 3-4, 135-142.
- [78] Lawson, J., Willden, C. Mixture Experiments in R Using mixexp. *Journal of Statistical Software, Code Snippets*, 2016, 72, 2, 1-20.
- [79] Levene, H. „Robust Tests for Equality of Variances.“ In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin et al, 1960, 278–292. Palo Alto, CA: Stanford University Press.
- [80] Lee, P. M. *Bayesian statistics: an introduction*. 4th ed., 2012, Hoboken, N.J.: Wiley.
- [81] Lin, D. K. J. A New Class of Supersaturated Designs, *Technometrics*, 1993, 35, 1, 28–31.
- [82] Lin, D. K. J. Generating systematic supersaturated designs. *Technometrics* 37, 1995, 213–225.

- [83] Long, J. S. Regression models for categorical and limited dependent variables, 1997, Thousand Oaks: Sage Publications.
- [84] Masri, M. A., Haberal, M., Rizvi, A., Stephan, A., Bilgin, N., Naqvi, A., Barbari, A., Kamel, G., N. Zafar, Emirog LU, R., Colak, T., Manzoor, K., Matha, V., Kamarád, V., Rost, M., Rizk, S., Hazime, A., Perlík, F. Switchability of Neoral and Equoral According to Food and Drug Administration Rules and Regulations. Transplantation Proceedings. New York : Elsevier Science Inc., 2005, 37, 2988-2993.
- [85] Masri, M., Rizk, S., Barbari, A., Stephan, A., Kamel, G., Rost, M. An assay for the determination of sirolimus levels in the lymphocyte of transplant patients. Transplantation Proceedings, 2007, 39, 1204-1206.
- [86] Masri, M., Rizk, S., Attia, M., Barbouch, H., Rost, M. Bioavailability of a New Generic Formulation of Mycophenolate Mofetil MMF 500 Versus CellCept in Healthy Adult Volunteers. Transplantation Proceedings. New York : Elsevier Science Inc., 2007, 39 (4), p. 1233-1236.
- [87] Másílko, J., Hartvich, P., Rost, M., Urbánek, M., Hlaváč, D., Dvořák, P. Potential for Improvement of Common Carp Production Efficiency by Mechanical Processing of Cereal Diet. Turkish Journal of Fisheries and Aquatic Sciences, 2014, 14:1, 145-153.
- [88] Mee, R. W. A comprehensive guide to factorial two-level experimentation, 2009, Dordrecht: Springer.
- [89] Moore, D. F. Applied survival analysis using R, 2016, New York, NY: Springer Science+Business Media.
- [90] Montgomery, D. C. Design and analysis of experiments. Eighth edition, 2013, Hoboken, NJ: John Wiley & Sons.
- [91] Morgan, J. M., Sonquist, J. A. Problems in the analysis of survey data, and a proposal, Journal of the American Statistical Association, 1963, 58, 415 - 434.
- [92] McLachlan G. J. Discriminant Analysis and Statistical Pattern Recognition, first ed., 1992, John Wiley & Sons, New Jersey.
- [93] McLean, R. A., Anderson, V. L. Extreme Vertices Design of Mixture Experiments, Technometrics, 1996, 8, 3, 447-454,



- [94] Mrkvička, T., Myllymäki M., Jílek M., Hahn, U. A one-way anova test for functional data with graphical interpretation, 2018, arXiv:1612.03608[stat.ME] dostupné na <https://arxiv.org/abs/1612.03608>
- [95] Mrkvička, T., Roskovec, T., Rost, M. A nonparametric graphical tests of significance in functional GLM, 2019, arXiv:1902.04926 [stat.ME], dostupné na <https://arxiv.org/abs/1902.04926>
- [96] Mullen, K., Ardia, D., Gil, D., Windover, D., Cline, J. 'DEoptim': An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software*, 2011, 40, 6, 1-26. URL <http://www.jstatsoft.org/v40/i06/>.
- [97] Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. and Hahn, U. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017, 79, 381-404.
- [98] Myers, R.H., Montgomery, D. C., Anderson-Cook, Ch. M. *Response surface methodology: process and product optimization using designed experiments*. 4th ed., 2016, Hoboken, New Jersey: Wiley.
- [99] Mynářová, A., Foitová, I., Kváč, M., Květoňová, D., Rost, M., Morrogh-Bernard, H., Nurcahyo, W., Nguyen, C., Supriyadi, S., Sak, B. Prevalence of *Cryptosporidium* spp., *Enterocytozoon bienersi*, *Encephalitozoon* spp. and *Giardia intestinalis* in Wild, Semi-Wild and Captive Orangutans (*Pongo abelii* and *Pongo pygmaeus*) on Sumatra and Borneo, Indonesia. *Plos ONE*, 2016, 11: 3,
- [100] Nash, J. C. *Compact numerical methods for computers: linear algebra and function minimisation* , 2nd ed, 1990, Bristol: Adam Hilger.
- [101] Nelder, J. A., Mead, R. A simplex algorithm for function minimization. *Computer Journal*, 1965, 7, 308–313.
- [102] Němejc, K., Sak, B., Květoňová, D., Kernerová, N., Rost, M., Cama, VA., Kváč, M. Occurrence of *Cryptosporidium suis* and *Cryptosporidium scrofarum* on commercial swine farms in the Czech Republic and its associations with age and husbandry practices *Parasitology Research*, 2013, 112:3, 1143-1154.
- [103] Nevill-Manning W. I. C., Holmes G. The development of holte's 1r classifier, in *Proceeding of the 2nd New Zealand Two-Stream Internati-*

- onal Conference on Artificial Neural Networks and Expert Systems, 1995, 63 - 91.
- [104] Nocedal, J., Wright, S. J. Numerical optimization, 2nd ed., 2006, New York: Springer.
- [105] Nyamundanda, G., Brennan, L., Gormley, I. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics*, 2010, 11, 1, 571.
- [106] Ondráčková, Z., Kváč, M., Sak, B., Květoňová, D., Rost, M. Prevalence and molecular characterization of *Cryptosporidium* spp. in dairy cattle in South Bohemia, the Czech Republic. *Veterinary Parasitology*. Elsevier B. V., 2009, 165, 1-5, 141-144.
- [107] Perlík, F., Masri, M. A., Rost, M., Kamarád, V. Pharmacokinetic conversion study of a new cyclosporine formulation in stable adult renal transplant recipients. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.* 2005, 149, 2, 309-13.
- [108] Plackett, R. L., Burman, J. P. The design of optimum multifactorial experiments, *Biometrika*, 1946, 33, 305–325.
- [109] Pinheiro, J. C., Bates, D. M. Mixed-effects models in S and S-PLUS, 2000, New York: Springer.
- [110] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1 - 117.
- [111] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [112] Thompson, L. A. R (and S-PLUS) Manual to Accompany Agresti's Categorical Data Analysis, 2009, 2nd edition.
- [113] Ramsay, J. O., Silverman, B. W., Applied functional data analysis: methods and case studies, 2002, New York: Springer. Springer series in statistics, ISBN 0-387-95414-7.
- [114] Ramsay, J. O., Hooker, G., Graves, S. Functional data analysis with R and MATLAB, 2009, New York: Springer.

- [115] Ripley, B. D. *Pattern Recognition and Neural Networks*, 1996, Cambridge University Press, Cambridge.
- [116] Rizzo, M. L., Székely, G. J. DISCO analysis: A nonparametric extension of analysis of variance. *The Annals of Applied Statistics*, 2010, 4: 2, 1034-1055.
- [117] Roy, R. K. *A primer on the Taguchi method*. 2nd ed. Dearborn, MI: Society of Manufacturing Engineers, 2010.
- [118] Russell V. L. Response-Surface Methods in R, Using rsm. *Journal of Statistical Software*, 2009, 32, 7, 1-17. URL <http://www.jstatsoft.org/v32/i07/>.
- [119] Rutkayová, J., Voříšková, J., Beneš, K., Hasoňová, L., Vácha, F., Křížek, M., Rost, M., Jirotková, D., Hanusová, L. The "Freshness" Objective Assessment of Carp Meat - Is It Appropriate to Use Freshness Meter? *Chemické listy*, 2018, 112, 5, 333-339.
- [120] Rutkayová, J., Vácha, F., Maršálek, M., Beneš, K., Civišová, H., Horká, P., Petrášková, E., Rost, M., Šulista, M. Fish stock losses due to extreme floods - findings from pond-based aquaculture in the Czech Republic. *Journal of Flood Risk Management*, 2018, 11, 3, 351-359.
- [121] Řezanková, H., Húsek, D., Snášel, V. *Shluková analýza dat*, 2. rozš. vyd., 2009, Praha: Professional Publishing.
- [122] Sak, B., Kváč, M., Květoňová, D., Kučerová, Z., Rost, M., Secor, E. Seropositivity for *Enterocytozoon bienersi*, Czech Republic. *Emerging Infectious Diseases*, 2010, 16, 2, 335-337.
- [123] Sak, B., Brady, D., Pelikánová, M., Květoňová, D., Rost, M., Kostka, M., Tolarová, V., Hůzová, Z., Kváč, M. Unapparent Microsporidial Infection among Immunocompetent Humans in the Czech Republic. *Journal of Clinical Microbiology*, 2011, 49, 3, 1064-1070.
- [124] Sak, B., Jandová, A., Doležal, K., Kváč, M., Květoňová, D., Hlásková, L., Rost, M., Olšanský, M., Nurcahyo, W., Foitová, I. Effects of selected Indonesian plant extracts on *E. cuniculi* infection in vivo. *Experimental Parasitology*, 2017, 181, 10, 94-101.
- [125] Scrucca, L. GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 2013, 53, 44, 1-37. URL <http://www.jstatsoft.org/v53/i04/>

- [126] Scrucca, L. On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution. *The R Journal*, 2017, 9, 1, 187-206. <https://journal.r-project.org/archive/2017/RJ-2017-008>
- [127] Schuirmann D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 1987, 15, 657–680.
- [128] Silverman, B. W. *Density Estimation for Statistics and Data Analysis*, 1993, Chapman and Hall, New York.
- [129] Slavíková, L., Malý, V., Rost, M., Petružela, L., Vojáček, O. Impacts of Climate Variables on Residential Water Consumption in the Czech Republic. *Water Resources Management*, 2013, 27, 2, 365-379.
- [130] Škuta, C., Svozil, D. QSAR - Modelování kvantitativních vztahů mezi strukturou a aktivitou chemických látek. *Chemické Listy*, 2017, 111, 747 - 753
- [131] Smith, W. F. S. *Experimental design for formulation*. Alexandria, 2005, Va.: American Statistical Association.
- [132] Suamant, P., Qasim, K. B., Rani, G. Optimization of alkaline protease from *Bacillus* sp. by response surface methodology, *Current Microbiology*, 2002, 44, 286 - 290.
- [133] Tipping, M. E., Bishop, C. M. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1999, 61, 3, 611–622.
- [134] Trygg, J., Wold, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, 2002, 16, 3, 119–128.
- [135] Urbánek, M., Hartvich, P., Vácha, F., Rost, M. Investigation of fat content in market common carp (*Cyprinus carpio*) flesh during the growing season. *Aquaculture nutrition*, 2010, 16, 5, 511-519.
- [136] Vácha, F., Cepák, M., Urbánek, M., Vejsada, P., Hartvich, P., Rost, M. Impact Of Long-Term Storage On The Instrumental Textural Properties Of Frozen Common Carp (*Cyprinus Carpio*, L.) Flesh. *International Journal of Food Properties*, 2013, 16, 2, 241-250.

- [137] Venables, W. N., Ripley B. D. Modern applied statistics with S-Plus, Fourth ed., 2002, Springer-Verlag, New York.
- [138] Vining, G. Technical Advice: Design of Experiments Response Surface Methodology, and Sequential Experimentation, Quality Engineering, Taylor & Francis Group, 2011, 23, 217 - 220.
- [139] Wadle, S., Lehnert, M., Rubenwolf, S., Zengerle, R., Stetten, F. Real-time PCR probe optimization using design of experiments approach, Biomolecular Detection and Quantification, 2016, 7, 1–8.
- [140] Wágnerová, P., Sak, B., McEvoy, J., Rost, M., Matysiak, A., Ježková, J., Kváč, M. Genetic diversity of *Cryptosporidium* spp. including novel identification of the *Cryptosporidium muris* and *Cryptosporidium tyzzeri* in horses in the Czech Republic and Poland. Parasitology Research, 2015, 114, 4, 1619-1624.
- [141] Wágnerová, P., Sak, B., McEvoy, J., Rost, M., Sherwood, D., Holcomb, K., Kváč, M. *Cryptosporidium parvum* and *Enterocytozoon bienersi* in American Mustangs and Chincoteague ponies. Experimental Parasitology, 2016, 162, 24-27.
- [142] Wand, M. P., Jones, M.C. Comparison of smoothing parameterizations in bivariate kernel density estimation, Journal of the American Statistical Association, 1993, 88, 520 - 528.
- [143] Wold, S., Sjostrom M., Eriksson, L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 2001, 58, 2, 109-130.
- [144] Wilcox, R. R. Introduction to robust estimation and hypothesis testing, 2012, 3rd ed. Boston: Academic Press.
- [145] Wu, Ch-F. Hamada, M. Experiments: planning, analysis, and optimization. 2nd ed., 2009, Hoboken, N.J.: Wiley, Wiley series in probability and statistics.
- [146] Yang, X.-S. Nature-inspired optimization algorithms, 2014, Boston: Elsevier.
- [147] Zhou, J., Xu, J.-J., Cong, J.-M., Cai, Z.-X., Zhang, J.-S., Wang, J.-L., Ren, Y.-P. Optimization for quick, easy, cheap, effective, rugged and safe extraction of mycotoxins and veterinary drugs by response surface

methodology for application to egg and milk, *Journal of Chromatography A*, 2018, 1532, 20-29.